Artigo

# REPOSITÓRIOS DIGITAIS CONFIÁVEIS E CERTIFICAÇÃO

Katia P. Thomaz
katia.thomaz@uol.com.br
Doutora em Ciência da Informação e bacharel em Ciência da Computação pela Universidade Federal de Minas Gerais - UFMG

**Resumo**

Sabe-se que para preservar nossos documentos e patrimônio digital nacional, precisaremos contar com uma rede de repositórios capazes de demonstrar confiabilidade para preservar esse conteúdo. E uma rede implica na capacidade de trabalharmos juntos, confiando uns nos outros – para trocar arquivos de dados e talvez se apoiar em certas instituições para preservar determinados conteúdos e outras instituições para outros. Com esses requisitos mínimos, começa-se a perceber uma nova necessidade no cenário da gestão e preservação de documentos digitais: a certificação formal de repositórios digitais. Essa questão foi tratada por um grupo de trabalho sobre atributos de arquivos digitais da junta RLG/OCLC, cujo relatório final Trusted Digital Repositories: Attributes and Responsibilities foi publicado em 2002. A partir desse relatório e novos trabalhos em andamento, este artigo aborda a questão e suas implicações para as organizações que pretendem preservar informação digital por longo prazo.

Palavras-chave: repositório digital confiável; sistema aberto de arquivamento de informação; certificação

**Abstract**

*We know that to preserve our nation's records and cultural heritage, we will need to have a networked group of repositories capable of being trusted to preserve that content. And being networked implies the ability to work together and to trust each other – to exchange files or perhaps instead rely on certain institutions to preserve some content while other institutions preserve others. With those minimal requirements, we begin to see that we have some firm needs emerging: the formal certification of digital repositories. All of these things were addressed by the RLG/OCLC Working Group on Digital Archive Attributes, and the final report Trusted Digital Repositories: Attributes and Responsibilities was issued in 2002. From RLG/OCLC report and new ongoing works, this paper discuss this subject and their implications for those organizations who intend to preserve digital information for long term.*

*Keywords: trusted digital repository; Open Archival Information System; certification*

## 1. INTRODUÇÃO

Sabe-se que para preservar nossos documentos e patrimônio digital nacional, precisaremos contar com uma rede de repositórios capazes de demonstrar confiabilidade para preservar esse conteúdo (THOMAZ e SOARES, 2004). E uma rede implica na capacidade de trabalharmos juntos, confiando uns nos outros – para trocar arquivos de dados e talvez se apoiar em certas instituições para preservar determinados conteúdos e outras instituições para outros. Com esses requisitos mínimos, começa-se a perceber uma nova necessidade no cenário da gestão e preservação de documentos digitais.

Essa questão foi tratada por um grupo de trabalho sobre atributos de arquivos digitais da junta RLG/OCLC1, cujo relatório final foi publicado em 2002. De maneira geral, esse relatório:

- propõe uma definição de repositório digital confiável,
- identifica os atributos primários de um repositório digital confiável,
- identifica as responsabilidades de um repositório compatível com o modelo de referência OAIS, e
- articula uma estrutura para o desenvolvimento de um programa de certificação.

A partir do relatório da RLG/OCLC (2002) e novos trabalhos em andamento, discutem-se três questões fundamentais relacionadas: confiança, modelo de referência SAAI (OAIS) e certificação.

## 2 CONFIANÇA

Sabe-se que confiança se desenvolve em diversos níveis. No caso de repositórios digitais confiáveis, no mínimo três níveis são aplicáveis:

- a confiança de que os produtores estão enviando as informações corretas,
- a confiança de que os consumidores estão recebendo as informações corretas, e
- a confiança de que os fornecedores estão prestando serviços adequados.

Mas como a "confiança" pode ser traduzida em um mecanismo mensurável? Como validamos a "confiança" que depositamos nas coisas? Porque as bibliotecas, os arquivos e os museus são encarregados de nosso patrimônio cultural nacional?

A resposta surge quase que naturalmente: Porque essas instituições adquiriram ao longo do tempo, a necessária confiança para armazenar esse material valioso. Essas instituições são confiáveis para fornecer acesso a esse material, com o objetivo de registrar e revelar a história, bem como fomentar o aumento do conhecimento. Elas são confiáveis para preservar esses itens nas melhores condições para futuras gerações.

---

[1] *Trusted Digital Repositories: Attributes and Responsibilities* em http://<www.rlg.org/longterm/repositories.pdf>.

Entretanto, essas instituições culturais têm sido bem sucedidas em preservar grandes quantidades de patrimônio cultural na forma de objetos físicos. Na maioria dos casos, esses objetos físicos estão disponíveis como "prova" da capacidade da instituição de recolher e preservar por longo prazo. Mas tendo em vista que a informação digital é menos tangível e muito mais mutável que outros materiais, confiança e credibilidade podem ser bem mais difíceis de comprovar...

Como mencionado anteriormente, o relatório da RLG/OCLC (2002) lista atributos e responsabilidades dos arquivos digitais confiáveis. Foram identificados os seguintes atributos:

- Conformidade com o modelo de referência SAAI
- Responsabilidade administrativa
- Viabilidade organizacional
- Sustentação financeira
- Adequação tecnológica
- Sistema de segurança
- Responsabilidade (accountability) de procedimentos

Observe-se que a adesão ao modelo de referência SAAI consta no topo dessa lista.

## 3 O MODELO DE REFERÊNCIA SAAI (OAIS)

O relatório da RLG/OCLC (2002) não descreve como o arquivo deve se apresentar ou se comportar, nem pretende estabelecer uma rede de repositórios idênticos. Mas se quisermos afirmar que um repositório digital está em conformidade com o modelo de referência SAAI, já saberemos algo sobre ele.

> Serviços de arquivamento digitais efetivos basear-se-ão no entendimento compartilhado de todo o conjunto de partes envolvidas, a respeito do que será atendido e como será atendido. O modelo de referência fornece uma estrutura comum, envolvendo terminologia e conceitos, para descrição e comparação de arquiteturas e operações de arquivos digitais. O SAAI ainda fornece tanto um modelo funcional – as atividades específicas realizadas pelo repositório como arquivamento ou acesso – quanto um modelo de informação correspondente que envolve um modelo para a criação de metadados para apoiar a manutenção e o acesso de longo prazo. Organizações e instituições que estabelecem repositórios digitais devem comprometer-se a entender esses modelos e esclarecer todos os aspectos da conformidade geral do sistema.[2] (RLG/OCLC WORKING GROUP ON DIGITAL ARCHIVE ATTRIBUTES, 2002)

Essa é a base da confiança a ser desenvolvida.

Recorde-se que um nível importante da confiança é a capacidade das instituições culturais confiarem nos serviços de terceiros. Se os serviços de terceiros forem compatíveis com o modelo de referência SAAI – ou apenas preencherem um dos aspectos de um sistema compatível com o modelo de referência SAAI –, ter-se-á o início de um entendimento compartilhado e um caminho mais fácil para o relacionamento confiável. Mas afinal, o que significa ser compatível com o modelo de referência SAAI? Qual o grau de esforço envolvido?

Segundo o modelo de referência SAAI (ABNT, 2007), a compatibilidade é atingida quando o arquivo atende aos modelos de informação e funcional propostos, mas principalmente quando cumpre um conjunto de responsabilidades, como a seguir:

- Negociar e aceitar informação adequada dos produtores de informação.
- Obter controle suficiente da informação fornecida no nível necessário para garantir a preservação por longo prazo.
- Determinar, por si mesmo ou em conjunto com outros parceiros, as comunidades que devem tornar comunidade alvo e, portanto, que devem ser capazes de entender a informação fornecida.
- Garantir que a informação a ser preservada seja independentemente compreensível para uma comunidade alvo. Em outras palavras, a comunidade alvo deve ser capaz de

[2] Para maiores informações sobre os modelos de informação e funcional, ver artigo da mesma autora e outro A preservação digital e o modelo de referência Opens Archival Information Systme (OAIS) – Parte 2, *DatagramaZero*, v.5, n.1, fev. 2004 em <http://www.dgz.org.br/fev04/F_I_art.htm>.

entender a informação sem a necessidade da assistência dos especialistas que produzem a informação.

- Seguir políticas e procedimentos documentados que garantam que a informação seja preservada contra todas as contingências cabíveis e que possibilitem que a informação seja disseminada como cópia autêntica do original ou rastreável até o original.
- Tornar a informação preservada disponível para a comunidade alvo (ABNT, 2007).

Como mencionado anteriormente, a confiança geralmente ocorre após um longo período de tempo. Entretanto, com a preservação digital, não temos tempo a perder. Precisamos de ações imediatas. Sendo assim, o que se pode fazer? Como tratar a preservação digital a curto-prazo para que possamos construí-la a longo-prazo?

## 4 CERTIFICAÇÃO

Certificação é a resposta.

> O modelo de referência SAAI não endereça diretamente a questão da certificação de arquivos. Entretanto, desperta um interesse considerável, especialmente quando um arquivo precisa basear-se em outro para um conjunto de serviços, sendo capaz de 'confiar' no outro arquivo. Uma forma de tratar essa questão de forma geral é identificar abordagens através das quais um arquivo possa estabelecer algum nível de certificação, seja por auto-avaliação ou por auditoria externa (RLG/OCLC WORKING GROUP ON DIGITAL ARCHIVE ATTRIBUTES, 2002)

A certificação tornou-se um componente-chave para repositórios digitais contemporâneos. Na dúvida, usamos a certificação como nosso mecanismo e instrumento de medida. Isso possibilita que repositórios se recuperem e caminhem, obtenham negócios, construam e comprovem boas práticas. E, ao longo do tempo, eles ganharão nossa CONFIANÇA. No passado, as práticas de certificação tendiam ao informal e implícito. Com os arquivos digitais há o desejo – talvez a necessidade – de tornar a certificação formalizada e explícita.

Em janeiro de 2007, o Consultative Committee of Space Data Systems - CCSDS reuniu um grupo de trabalho (Birds of Feather – BOF)[3] para tratar o assunto e, aos moldes do modelo de referência SAAI, desenvolver uma norma internacional para auditoria e certificação de repositórios digitais.

O plano de trabalho prevê a apresentação de um esboço em março de 2008, de um "livro branco" [terminologia adotada no âmbito do CCSDS] em outubro de 2008 e de um "livro vermelho", incluindo a norma ISO, em março de 2009. As principais fontes de referência desse trabalho têm sido as seguintes:

[3] http://wiki.digitalrepositoryauditandcertifiation.org/

Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC) por RLG/NARA Task Force on Digital Repository Certification[4];

Catalogue of Criteria for Trusted Digital Repositories por nestor[5] Working Group on Trusted Repositories Certification[6];

Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) por DCC/DPE[7]; e

Extratos da ISO/IEC 27001 Information Technology – Security techiniques – Information security management systems – Requirements no que diz respeito à segurança.

Uma análise inicial dessas fontes de referência revela os aspectos centrais da certificação de repositórios digitais.

## 5 OS PILARES DA CERTIFICAÇÃO

Três aspectos ou pilares centrais da certificação foram identificados nesse estudo das fontes de referência: organização, tecnologia e gerenciamento. A FIG. 1 ilustra essa tríade.


**FIGURA 1: Três pilares da certificação de repositórios digitais**

[4] http://www.crl.edu/PDF/trac.pdf
[5] Acrônimo de Network of Expertise in long-term STORage of Digital Resources. Para informações sobre o projeto, consultar http://www.langzeitarchivierung.de
[6] http://edoc.hu-berlin.de/series/nestor-materialien/8en/PDF/8en.pdf
[7] http://www.repositoryaudit.eu/

O pilar organizacional certifica o encargo, o escopo, os objetivos, a disponibilidade financeira e os compromissos de uma organização para se engajar na preservação digital. Esses compromissos devem envolver especialmente a adesão a normas prevalecentes sobre gestão e preservação digital, como as normas ISO 15489 Information and documentation - Records management e NBR 15472 Sistemas espaciais de dados e informações - Modelo de referência para um sistema aberto de arquivamento de informação (SAAI).

Podem contribuir para o atendimento aos requisitos organizacionais, as seguintes iniciativas:

estudos sobre custos da preservação digital, como os modelos econômicos propostos por Brian Lavoie[8];

pesquisa sobre modelos e estruturas de custos de Shelby Sanett[9];

abordagem desenvolvida pela Biblioteca Nacional da Holanda que estabelece uma ferramenta para comparar custos de migração e emulação ao longo do tempo[10]; e

fórmula de custo mais abrangente proposta pelo projeto Life Cycle Information for E-Literature - LIFE[11].

Os exemplos anteriores podem levar ao maior entendimento sobre os custos envolvidos na preservação digital dentro de determinada comunidade, mas não se equiparam a uma documentação abrangente da própria comunidade sobre as questões organizacionais.

O pilar tecnológico certifica a adequação da infra-estrutura técnica do repositório e sua capacidade de atender demandas de gerenciamento e segurança do repositório e dos objetos digitais. Ele envolve hardware, software, mídias de armazenamento, redes, medidas de segurança, gerenciadores de fluxo de trabalho (workflows), protocolos, documentação e habilidades tanto técnicas quanto administrativas.

Podem contribuir para o atendimento aos requisitos tecnológicos as seguintes iniciativas:

certificações em sistemas de gerenciamento da segurança da informação, como a norma ISO 27001 Information technology - Security techniques - Information security management systems - Requirements, e

aplicativos de repositório, como DSpace[12], Flexible Extensible Digital Object and Repository Architecture - Fedora[13] e Dark Archive In The Sunshine State - DAITSS[14].

[8] http://www.oclc.org/research/projects/digipres/economics.htm
[9] http://www.rlg.org/preserv/diginews/diginews7-4.html#feature2
[10] http://www.rlg.org/en/page.php?Page_ID=20571#article0
[11] http://www.ucl.ac.uk/ls/life
[12] http://www.dspace.org
[13] http://www.fedora.info
[14] http://daitss.fcla.edu

Mas é bom lembrar que, mesmo com esses exemplos de aplicativos disponíveis, as organizações precisam decidir como selecionar a opção mais adequada, considerando suas capacidades e limitações e a extensão na qual o aplicativo de repositório atende aos requisitos arquivísticos e se adapta ao conteúdo digital a ser preservado.

As organizações também podem optar pelo desenvolvimento de seu próprio repositório, como o Arquivo Nacional do Reino Unido[15], ou contratar um provedor de serviço de preservação digital, como Berkeley Electronic Press - bepress[16] ou o Arquivo Digital da OCLC[17].

O pilar gerencial certifica as funções, processos e procedimentos necessários para gerenciar os objetos digitais propriamente ditos, agrupados segundo as entidades funcionais definidas no famoso modelo de referência SAAI.

Podem contribuir para o atendimento aos requisitos gerenciais as seguintes iniciativas:

normas que detalham o modelo de referência SAAI, como a norma ISO 20652 Space data and information transfer systems - Producer-archive interface - Methodology abstract standard, já com sua adaptação para NBR em andamento no Comitê Brasileiro Aeronáutica e Espaço – CB08 da Associação Brasileira de Normas Técnicas - ABNT;

normas de requisitos, como o e-ARQ[18] do Arquivo Nacional do Brasil;

normas de metadados, como a norma ISO 23081 Information and documentation - Records management processes - Metadata for records, o Metadata Encoding & Transmission Standard - METS[19] da Biblioteca do Congresso norte-americano e o dicionário de dados PREMIS[20] da OCLC/RLG; e

ferramentas de preservação digital em determinados pontos do fluxo de trabalho.

Exemplos recentes de ferramenta de preservação digital incluem:

aquelas que recolhem material Web, p.ex., o Web Curator Tool - WCT[21] fruto do trabalho conjunto entre a Biblioteca Nacional da Nova Zelândia e a Biblioteca Nacional Britânica;

aquelas que documentam formatos de arquivos de dados, p.ex., o sítio de formatos[22] da Biblioteca do Congresso Nacional norte-americano e o PRONOM[23] do Arquivo Nacional do Reino Unido;

---

[15] http://www.nationalarchives.gov.uk/preservation/digitalarchive/default.htm
[16] http://www.bepress.com/
[17] http://www.oclc.org/digitalarchive/defaul.htm
[18] http://www.conarq.arquivonacional.gov.br/Media/publicacoes/earqbrasilv1.pdf
[19] www.loc.gov/standards/mets/
[20] Acrônimo de Preservation Metadata: Implementation Strategies. Para maiores informações sobre o dicionário de dados, consultar http://www.loc.gov/standards/premis/
[21] http://webcurator.sourceforge.net

aquelas que identificam e avaliam formatos de arquivos de dados, p.ex., o JSTOR/Harvard Object Validation Environment - JHOVE[24] e o Digital Record Object Identification - DROID[25];

aquelas que normalizam arquivos de dados para formatos preserváveis, p.ex., o XML Electronic Normalising of Archives - XENA[26];

aquelas que geram e capturam metadados, p.ex., o extrator de metadados[27] da Biblioteca Nacional da Nova Zelândia; e/ou

aquelas que produzem um identificar único e auxiliam na detecção de mudanças em arquivos de dados (p.ex., soma de fechamento).

Aqui também é bom destacar que, mesmo com essas ferramentas disponíveis, a comunidade arquivística tem muitos caminhos a percorrer antes que a preservação digital seja totalmente automatizada e que sistemas de preservação digital totalmente compatíveis estejam disponíveis.

## 6 CONSIDERAÇÕES FINAIS

Um repositório digital confiável é mais do que uma organização encarregada de armazenar e administrar objetos digitais. Um repositório digital confiável é "aquele cuja missão é fornecer acesso confiável, por longo prazo, a recursos digitais administrados à sua comunidade-alvo, agora e no futuro" (RLG/OCLC WORKING GROUP ON DIGITAL ARCHIVE ATTRIBUTES, 2002). Algumas instituições poderão implementar todas as responsabilidades funcionais e de informação em um sistema local completo. Outras preferirão administrar os aspectos lógicos e intelectuais de um repositório, contratando terceiros para os serviços de armazenamento e de manutenção dos objetos digitais. Independentemente, o componente-chave será a infra-estrutura geral para apoiar a confiabilidade e a sustentabilidade de repositórios digitais, de forma que as organizações e suas comunidades-alvo possam confiar que os recursos digitais serão preservados por longo prazo.

Para as organizações que tencionam fornecer serviços de repositório digital, o desenvolvimento da desejável confiança via práticas confiáveis, comprovadas, levará algum tempo. Entretanto, tendo em vista que ações imediatas para preservar o já extenso corpo de materiais digitais precisam ser tomadas, um programa de certificação seria recomendável para fornecer uma base de confiança. A certificação especificaria os critérios a serem atingidos e

[22] http://www.digitalpreservation.gov/formats/index.shtml
[23] http://www.nationalarchives.gov.uk/pronom/
[24] http://hul.harvard.edu/jhove/
[25] http://droid.sourceforge.net/wiki/index.php/Introduction
[26] http://xena.sourceforge.net/index.html

empregaria mecanismos para sua avaliação e medição. As instituições culturais contariam com uma ferramenta para medição contínua dos serviços disponíveis e os fornecedores de serviços teriam um conjunto conhecido de melhores práticas ou normas a atender, de forma a obter negócios dessas instituições culturais. A certificação, periodicamente atendida ao longo de diversos anos, poderia solucionar a tensão entre a necessidade imediata de arquivos confiáveis e a necessidade de desenvolver e comprovar a confiabilidade ao longo do tempo.

Finalmente, é preciso ter em mente que o movimento rumo a uma norma internacional para auditoria e certificação de repositórios digitais tem o referendo de duas conceituadas instituições de âmbito internacional com larga experiência no desenvolvimento de padrões, CCSDS e ISO, além de surgir do já amplamente adotado modelo de referência para um sistema aberto de arquivamento de informação – SAAI, o qual precisa ser imediatamente entendido e aplicado[28].

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS (ABNT). *NBR 15472*: Sistemas espaciais de dados e informações - Modelo de referência para um sistema aberto de arquivamento de informação (SAAI). 2007.

Digital Curation Centre (DCC); DigitalPreservationEurope (DPE). *Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)*. 28 Feb. 2007. Disponível em: <http://www.repositoryaudit.eu/download/>. Acesso em: 23 julho 2007.

INTERNATIONAL ORGANIZATION OF STANDARDIZATION (ISO). *ISO/IEC 27001*: Information technology: Security techniques: Information security management systems: Requirements. 2005.

MCGOVERN, Nancy 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://www.rlg.org/en/page.php?Page_ID=21033#article3>. Acesso em: 23 julho 2007.

NESTOR WORKING GROUP ON TRUSTED REPOSITIORIES CERTIFICATION. *Catalogue of Criteria for Trusted Digital Repositories*. Dec. 2006. Disponível em: <http://edoc.hu-berlin.de/series/nestor-materialien/8en/PDF/8en.pdf>. Acesso em: 23 julho 2007.

RLG/NARA TASK FORCE ON DIGITAL REPOSITORY CERTIFICATION. *Trustworthy repositories audit & certification*. Feb. 2007. Disponível em: <http://www.crl.edu/PDF/trac.pdf >. Acesso em: 23 julho 2007.

RLG/OCLC WORKING GROUP ON DIGITAL ARCHIVE ATTRIBUTES. *Trusted digital repositories: attributes and responsibilities*. May 2002. Disponível em: <http://www.rlg.org/longterm/repositories.pdf>. Acesso em: 23 julho 2007.

THOMAZ, Katia P.; SOARES, Antonio José. A preservação digital e o modelo de referência Open Archival Information System (OAIS). *DataGramaZero*, v. 5, n. 1, fev. 2004. Disponível em:<http://www.dgz.org.br/fev04/F_I_art.htm>. Acesso em: 23 julho 2007.

[27] http://meta-extractor.sourceforge.net

[28] Para acompanhar os trabalhos de desenvolvimento da norma internacional para auditoria e certificação de repositórios digitais, consultar http://wiki.digitalrepositoryauditandcertification.org/.