

# UMA OUTRA FACE DOS METADADOS: INFORMAÇÕES PARA A GESTÃO DA PRESERVAÇÃO DIGITAL

Luís Fernando Sayão  
[lsayao@cnen.gov.br](mailto:lsayao@cnen.gov.br)  
CNEN - Rio de Janeiro - RJ

## Resumo

O conceito tradicional de metadado pode ser ampliado para abrigar um conjunto de informações que apoiem as atividades de gestão da preservação de materiais digitais. Esse tipo de metadados, chamados de metadados de preservação, tem como função instruir e documentar os processos de preservação digital de longo prazo, garantindo que os conteúdos digitais possam ser acessados e interpretados no futuro. Nos últimos anos, foram desenvolvidos inúmeros esquemas e infraestruturas de metadados voltados para a preservação digital, que tiveram como maior desafio antecipar quais informações são realmente necessárias para suportar um processo específico de preservação. A iniciativa mais importante e mais abrangente nesse campo é o dicionário de dados PREMIS, cujo desenvolvimento teve como base a infraestrutura conceitual definida pela norma OAIS. A ideia básica deste trabalho é revisar os principais conceitos, padrões e tecnologias envolvidos no desenvolvimento de esquemas de metadados de preservação.

**Palavras-chave:** Metadados para preservação. Preservação digital. PREMIS. OAIS. METS.



Esta obra está licenciada sob uma [Licença Creative Commons](https://creativecommons.org/licenses/by/4.0/).

DOI 10.5007/1518-2924.2010v15n30p1

# 1 INTRODUÇÃO

Nos dias de hoje, é virtualmente impossível discutir serviços e sistemas de informação sem o envolvimento direto com questões relacionadas aos metadados. Embora o termo “metadados” seja uma invenção relativamente recente – primordialmente ele foi usado no contexto dos sistemas de banco de dados para descrever e controlar a gestão e o uso dos dados - a ideia que ele porta remonta outros tempos, tendo suas raízes na catalogação realizada pelas bibliotecas e organizações similares (DAY, 2005). Essa noção é determinante, posto que, quando pensamos em metadados, a primeira idéia que nos ocorre é inspirada no seu uso no ambiente da biblioteca; no seu papel de um esquema formal para descrição de todo tipo de objetos informacionais, digitais e não digitais. A catalogação tradicional é uma forma de atribuição de metadados; o MARC 21<sup>1</sup> e o conjunto de regras usadas com ele, tais como o AACR2<sup>2</sup>, são padrões de metadados (NISO, 2004).

Relacionada à função de catalogação, existe outra importante razão para a criação de metadados: facilitar a descoberta de informações relevantes, seja no ambiente da biblioteca, seja no ambiente *web*. O exemplo mais ilustrativo é o Dublin Core Metadata Element Set<sup>3</sup>, uma das mais importantes iniciativas na área de metadados, cujo objetivo essencial é apoiar a descoberta de recursos no extenso e fragmentado universo *web*, que apesar da sua riqueza informacional não foi pensado especificamente para a recuperação de informação.

Porém, quando uma biblioteca assinala metadados descritivos para um livro de sua coleção, ela não precisa se preocupar com a possibilidade dele se dissolver numa série de páginas e figuras desconectadas caso as informações sobre a seqüência das páginas e a estrutura do livro não forem registradas; nenhum pesquisador ficará impossibilitado de avaliar o conteúdo do livro se os dados sobre a máquina *offset* que o imprimiu não forem informados. O mesmo não pode ser dito para a versão digital desse livro (LIBRARY OF CONGRESS, 2009). Quando submergimos no mundo dos documentos digitais, constatamos que outras dimensões dos metadados, que ultrapassam os limites de ferramenta para a descrição e descoberta de recursos, precisam ser reveladas e

---

<sup>1</sup> <<http://www.loc.gov/marc/>>

<sup>2</sup> <<http://www.aacr2.org/>>

<sup>3</sup> <<http://dublincore.org/>>

exploradas. Isto porque os objetos digitais para serem gerenciados e usados requerem processos de maior amplitude, que implica em identificar informações precisas para instruí-los adequadamente.

Na medida em que a idéia de metadados se torna uma parte essencial do mundo digital, eles se mostram conceitualmente mais complexos e mais abrangentes, apoiando um espectro extremamente amplo de atividades. Essas novas dimensões de metadados são vitais para o acesso e para a interpretação dos recursos informacionais digitais; como são importantes também para a estruturação e para os processos de gestão associados a esses recursos, que podem incluir inúmeras funções, tais como: controle dos direitos, intercâmbio, comércio eletrônico, interoperabilidade técnica e semântica, reuso da informação e curadoria digital, para citar alguns. Esse elenco crescente de funções circunscreve conceitos tradicionais e conceitos inéditos que convergem para apoiar a composição de novos ambientes informacionais, como as bibliotecas, os arquivos e os museus digitais.

Esta ampliação do domínio de aplicação faz com que os metadados necessários para a gestão e para o uso de objetos digitais sejam mais diversificados e, na maioria dos casos, diferentes dos metadados usados para gestão de coleções de obras impressas e de outros materiais físicos.

Em outro plano, o acesso e a usabilidade dos recursos informacionais digitais é impactado fortemente pela sua dependência a contextos tecnológicos específicos; esse fato gera uma área de tensão e complexidade na gestão de acervos digitais. A fragilidade estrutural da informação digital configura um dos maiores desafios a ser enfrentado pelos pesquisadores e profissionais das áreas de informação e de tantas outras áreas, neste começo de século. A preservação da informação digital por longo prazo é um problema que envolve um número grande de variáveis, planejamento cuidadoso, tecnologia e orçamentos vultosos, e cuja complexidade tem arrefecido o entusiasmo das bibliotecas digitais e demais organizações de patrimônio informacional em disponibilizar seus estoques digitais para as futuras gerações.

Entretanto, está cada vez mais claro – para a prática e para a teoria – que existe uma parte do problema de preservação digital de longo prazo que só será resolvido a partir da identificação de um conjunto de dados e informações, expressos na forma de metadados, que ancorem os processos de gestão da preservação digital.

Este elenco específico de metadados é chamado de metadados de preservação; é uma nova face para os metadados que vai assegurar que o recurso de valor contínuo sobreviva ao longo do tempo e continue sendo acessível e, não menos importante, que não perca a capacidade de ter seus significados apropriadamente interpretados no tempo que for necessário pelas comunidades para quem a informação, de forma privilegiada, se dirige.

Nessa direção, uma série de especificação de metadados e de infraestruturas físicas e conceituais vem sendo desenvolvida em torno do compromisso da preservação de longo prazo das informações digitais.

É exatamente o papel dos metadados como ferramenta voltada para instruir os processos de preservação de documentos digitais que vamos discutir resumidamente nesse trabalho. Para contextualizar o problema, começamos com uma rápida definição de metadados e seus tipos; passamos pelas estratégias de preservação digital; em seguida, discutimos os metadados de preservação, tomando como referências o modelo conceitual definido pelo Open Archival Information System (OAIS), o dicionário de dados do PREMIS<sup>4</sup> e o papel da infraestrutura de empacotamento definida pelo Metadata Encoding Transmission Protocol (METS)<sup>5</sup>.

## **2 UMA DEFINIÇÃO E UMA CATEGORIZAÇÃO PARA METADADOS**

Primordialmente, as iniciativas relacionadas à criação de formatos de metadados estavam focadas no desenvolvimento de padrões para organização e para a descoberta de recursos informacionais. Entretanto, novas exigências, impostas principalmente pelos desafios do mundo digital, foram redesenhando a ideia puramente descritiva de metadados, criando expansões para o seu conceito com o intuito de abrigar novos propósitos e funções.

Como desdobramento, a definição minimalista e quase clássica, que enuncia que “metadados é dados sobre dados”, torna-se inexpressiva e rasa diante da complexidade dos papéis atribuídos aos metadados nos diversos contextos correntes da gestão da informação; além do mais, ela não

---

<sup>4</sup> <<http://www.oclc.org/research/projects/pmwg/default.htm>>

<sup>5</sup> <<http://www.loc.gov/standards/mets/>>

nos ajuda a entender o que é e como os metadados podem ser usados. A NISO<sup>6</sup> – sigla para National Information Standard Organization - apresenta uma definição que expande o que se entende por metadados, ampliando o seu domínio de aplicação: “Metadados é a informação estruturada que descreve, explica, localiza, ou possibilita que um recurso informacional seja fácil de recuperar, usar ou gerenciar” (NISO, 2004, p.1, tradução nossa).

Não se pode afirmar que haja um consenso, mas uma fração significativa dos autores que tratam do assunto concorda que os metadados podem ser divididos em três categorias conceituais: metadados descritivos, metadados estruturais e metadados administrativos. Essa segmentação é útil para uma compreensão mais clara sobre os tipos de informações que eles podem circunscrever, muito embora os seus contornos não possam ser precisamente definidos.

- **Metadados descritivos:** é a face mais conhecida dos metadados, são eles que descrevem um recurso com o propósito de descoberta e identificação; podem incluir elementos tais como título, autor, resumo, palavras-chave e identificador persistente.
- **Metadados estruturais:** são informações que documentam como os recursos complexos, compostos por vários elementos, devem ser recompostos e ordenados. Por exemplo, como as páginas de um livro, digitalizadas separadamente, são vinculadas entre si e ordenadas para formar um capítulo.
- **Metadados administrativos:** fornecem informações que apoiam os processos de gestão do ciclo de vida dos recursos informacionais. Incluem, por exemplo, informações sobre como e quando o recurso foi criado e a razão da sua criação. Nessa categoria, estão metadados técnicos que explicitam as especificidades e dependências técnicas do recurso; inclui também os metadados voltados para apoio à gestão dos direitos relacionados ao recurso.

Um requisito importante para os sistemas de informações atuais é a possibilidade da representação de recursos informacionais em níveis variados de granularidade; isso compreende a capacidade dos metadados de descreverem camadas diferenciadas de agregação dos recursos, por exemplo: descrever uma coleção, um item ou uma parte de um item, como um capítulo, uma fotografia ou um gráfico. Ainda relacionada à amplitude de resolução dos metadados, está a

---

<sup>6</sup> <<http://www.niso.org/home>>

capacidade de descrever uma obra e suas expressões, manifestações e itens particulares (NISO, 2004).

Metadados são agrupados em estruturas abstratas conhecidas como esquemas ou formatos de metadados, que são conjuntos de elementos criados com fins específicos, por exemplo: descrever um tipo particular de recurso de informação. Muitos e diferentes esquemas de metadados têm sido continuamente desenvolvidos tendo como perspectiva uma grande variedade de usos em contextos variados, porém cada qual é limitado por suas especificidades e pelos seus domínios de aplicação próprios. Os poucos exemplos a seguir nos mostram um pouco dessa diversidade: MODS (Metadata Object Description Schema)<sup>7</sup> esquema bibliográfico derivado do MARC 21; EAD (Encoded Archival Description)<sup>8</sup> voltado para a área de Arquivologia; LOM (Learning Object Metadata)<sup>9</sup> para gerenciar, avaliar e localizar objetos de aprendizagem; MPEG Multimedia Metadata<sup>10</sup> para representação de objetos multimídiaicos.

O esquema de metadados Dublin Core, por sua vez, cria uma situação especial, posto que não está focado em nenhum tipo específico de objeto ou de domínio de assunto; está voltado para descoberta de recursos em domínios transversais; e é minimalista por natureza, sendo composto por poucos elementos essenciais (o *core*), passíveis de serem mapeáveis em outros formatos, constituindo a *lingua franca* dos metadados e uma das chaves para o santo graal da interoperabilidade. Outra característica importante do Dublin Core é ser auto-explicativo o suficiente para permitir que o próprio autor – ou melhor, criador – da obra possa descrevê-la e publicá-la na *web*. Não obstante, o esquema possui uma estrutura simples e flexível e pode ser aplicado a recursos complexos; além do mais, pode ser representado através de sintaxes variadas, por exemplo, codificado em HTML ou em XML e estruturado segundo a arquitetura proposta pela RDF (Resource Description Framework)<sup>11</sup>, facilitando o intercâmbio e o reuso.

É importante notar que os metadados podem estar embutidos num objeto digital inscrito na sua codificação, como é comum nos documentos HTML e XML ou no *header* de arquivos de imagens; ou podem estar armazenados separadamente, estruturados em bases de dados, facilitando a busca e a recuperação, como num catálogo *on-line* no ambiente de biblioteca. No

---

<sup>7</sup> <<http://www.loc.gov/standards/mods/>>

<sup>8</sup> <<http://www.loc.gov/ead/>>

<sup>9</sup> <<http://ltsc.ieee.org/wg12/>>

<sup>10</sup> <<http://www.chiariiglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>>

<sup>11</sup> <<http://www.w3.org/RDF/>>

mundo da *web*, os metadados precisam também ser compreendidos por computadores, por meio de robôs e agentes de software, para que possam ser recuperados e tenham sua relevância avaliada e sejam manipulados com maior eficiência. O uso de programas para processar metadados codificados em XML é um dos pilares da iniciativa denominada *web semântica*. (MARCONDES, 2005).

### 3 PERSERVAÇÃO DIGITAL E O PAPEL DOS METADADOS

O artefato digital traz consigo uma fragilidade estrutural intrínseca que coloca permanentemente em risco a sua longevidade, tornando a preservação dos conteúdos em formatos digitais um dos desafios essenciais do nosso tempo. O problema da instabilidade das informações digitais, que nos ameaça com uma espécie de amnésia digital e uma nova pré-história, está inscrito na agenda crítica da humanidade, acompanhando outros desdobramentos negativos da tecnologia, a espera de uma solução completa e abrangente (CONARQ, 2004).

A preservação digital, enquanto um conjunto de atividades voltadas para garantir o acesso aos conteúdos digitais por longo prazo, é, ao mesmo tempo, um desafio técnico e organizacional que se desenrola permanentemente no tempo e no espaço; seus objetivos exigem processos que portem uma intencionalidade contínua, dado que os objetos digitais não sobrevivem inercialmente, como sobrevivem as plaquetas de argila de cinco mil anos encontradas casualmente no deserto. Não existe absolutamente essa possibilidade para os objetos digitais.

As ameaças que cercam os objetos digitais são engendradas pela sua própria condição física, não fixada em suportes e fortemente dependente de contextos tecnológicos específicos e fugazes. Pela primeira vez na história, temos que preservar registros que não estão ao alcance de nenhum dos nossos sentidos, como os papiros egípcios e os pergaminhos romanos, registros cuja materialidade estruturada em átomos e moléculas está mais evidente do que os padrões virtuais - formados por *bits* e *bytes* que estabelecem a *fisicalidade* dos objetos digitais.

Ao contrário de uma carta ou de um livro impresso, em que a leitura e a interpretação são ações diretas e sem intermediação, entre um objeto digital e seu usuário se interpõe um ambiente tecnológico complexo e específico, formado por camadas de *software* (sistema operacional,

aplicativos, etc.), *hardware*, tecnologia de redes e equipamentos especiais. “Por esta razão, não basta simplesmente preservar o objeto digital: os meios de apresentar e de usar o objeto devem também ser preservados” (LAVOIE; GARTNER, 2005, p.6). Isso implica ter disponível, para acesso aos conteúdos e às funcionalidades do objeto digital, o ambiente correto ou, pelo menos, um substituto tecnologicamente equivalente.

Entretanto, esse ambiente tecnológico, insuflado pela inovação, competitividade e mercados em expansão, tem um ciclo de evolução continuamente mais dinâmico, tornando-se ultrapassado em lapsos de tempo cada vez menores; esse fato coloca como imprescindível que se documente cuidadosamente o ambiente tecnológico necessário para acesso e uso dos objetos digitais arquivados.

Outra característica crítica dos objetos digitais é que eles são altamente suscetíveis a alterações (intencionais ou não) e à fragilidade das mídias, cuja gradual degradação pode levar a perdas parciais ou totais de informações. A mutabilidade dos objetos digitais tem impacto significativo na fixação e na manutenção de sua aparência e da sua usabilidade; mesmo as ações de preservação podem alterar a forma e a função de um objeto digital. Essa transitoriedade dos objetos digitais torna essencial que eles estejam acompanhados de informações que documentem as suas características, sua história, incluindo todas as alterações sofridas por eles.

Por fim, é necessário considerar que operações sobre objetos digitais podem estar limitadas por cláusulas de direitos de propriedade intelectual, que podem impor limitações às ações de preservação digital, posto que, em muitos casos, elas implicam em intervenções sobre o conteúdo, funcionalidades e aparência dos objetos. Por esse motivo, é necessário documentar os direitos associados aos objetos arquivados, para que os processos de preservação estejam coordenados com as restrições impostas aos objetos (LAVOIE; GARTNER, 2005).

Desde os primeiros momentos da criação de dados e informações em meio eletrônico, já se previa que estes problemas seriam os *leviatãs* que ameaçariam o acesso persistente aos conteúdos digitais e trariam a incerteza de que a aparência, as funcionalidades, a autenticidade e a integridade desses conteúdos poderiam não ser recompostas no futuro. Entretanto, essa preocupação vem se tornando dramaticamente mais crítica, na medida em que segmentos importantes da sociedade moderna – a pesquisa científica, o governo, os negócios, a cultura e a educação - dependem mais e mais de informações digitais, na maioria das vezes produzidas por

eles mesmos, como elemento essencial para todos os seus empreendimentos; e que patrimônios digitais valiosos já foram perdidos para sempre, por exemplo, parte significativa das informações sobre a exploração do planeta Marte pela sonda americana *Viking* na década de 1970 (BESSER, 1999) e as primeiras mensagens de correio eletrônico trocadas entre os cientistas na década de 1960 (LUKESH, 1999), testemunhos do início de uma época que, ironicamente, não sobreviveram à própria essência desse tempo, a transitoriedade da tecnologia.

### 3.1 Estratégias de preservação digital

As funções de preservação podem variar de repositório para repositório, mas geralmente circunscrevem ações que asseguram que os objetos digitais permaneçam *viáveis*, isto é, que possam ser lidos a partir de uma mídia; que possam ser *apresentados*, ou seja, possam ser visualizados, executados ou interpretados pelo software de aplicação; e que mantenham sua *integridade*, significando não serem alterados inadvertidamente e que as mudanças legítimas sofridas tenham sido documentadas (CAPLAN, 2009).

As estratégias de preservação digital que estão sendo praticadas e pesquisadas pelas comunidades envolvidas com o problema de acesso, a longo prazo, a informações digitais são resumidas a seguir<sup>12</sup>:

- **Preservação da tecnologia** – estratégia baseada na criação de museus tecnológicos que mantêm equipamentos e software obsoletos, de forma que os documentos digitais possam ser processados no seu ambiente original. É uma solução de curto prazo.
- **Emulação** – estratégia fundamentada na premissa de que o melhor meio de preservar as funcionalidades e a aparência de um objeto informacional digital é preservá-lo junto ao seu software original; dessa forma, o objeto pode ser rodado em plataformas atuais por meio de emuladores, que são programas que criam mímicas do comportamento de *hardware* e sistemas operacionais obsoletos em

---

<sup>12</sup> Para uma análise mais aprofundada, recomenda-se o estado da arte publicado por Lee e seus colaboradores ainda em 2002, mas que permanece atual.

computadores novos. Essa estratégia tem sido foco de muitas pesquisas e controvérsias.

- **Migração** – tem como fundamento a migração periódica de um patamar tecnológico em vias de se tornar obsoleto e/ou de se degradar fisicamente para outro mais atualizado e íntegro, incluindo mídias, ambientes de software, formatos e computadores; é a estratégia correntemente mais utilizada pelas organizações (SAYÃO, 2007).
- **Encapsulamento** – baseia-se na idéia de que os objetos preservados devem ser autodescritos e encapsulados em estruturas físicas ou lógicas com todas as informações necessárias para que seja decifrado e compreendido no futuro.

### 3.2 Metadados de Preservação

Todas essas estratégias, para alcançarem seus objetivos, dependem fortemente da captura, criação e manutenção de vários tipos de dados que informem sobre histórico, características técnicas, estruturas, dependências e alterações sofridas pelo objeto digital. São esses dados que viabilizarão o pleno acesso e permitirão a recriação e a interpretação da estrutura e do conteúdo da informação digital ao longo do tempo. Para tal, eles são estruturados na forma de metadados, compondo o que chamamos de “metadados de preservação”.

Dessa forma, os metadados de preservação constituem uma parte essencial das estratégias de preservação digital. A síntese de sua importância pode ser expressa pelo fato deles permitirem que um objeto digital esteja autodocumentado ao longo do tempo e, portanto, posicionado para a preservação de longo prazo e para o acesso contínuo, apesar da sua propriedade, custódia, tecnologia, restrições legais, e mesmo da sua comunidade de usuários estar continuamente mudando (LAVOIE; GARTNER, 2005, P.7).

Os metadados de preservação podem ser definidos, de uma forma simples e direta, como a informação que apoia e documenta a preservação de longo prazo de materiais digitais. Entretanto, com o provável intuito de se alinhar ao consenso de que a preservação digital é um processo de

gestão, alguns autores categorizam os metadados de preservação como metadados administrativos. Porém, com um grau a mais de aproximação, verificamos que os esquemas de metadados de preservação incluem elementos que se enquadram em todas as três categorias – descritivos, administrativos e estruturais. Considerando essa maior abrangência, podemos reescrever a definição de metadados de preservação mais precisamente como “metadados descritivos, estruturais e administrativos que apoiam e documentam a preservação de longo prazo de materiais digitais” (DAY, 2003).

Definidos dessa forma, fica claro que os metadados de preservação são criados para apoiar um grande número de funções diferentes, porém relacionadas. O amplo espectro de funções, que se espera que os metadados de preservação cumpram, sinaliza que a definição de um padrão é uma tarefa difícil e de grande amplitude; a maioria dos esquemas atualmente publicados é extremamente complexa ou somente estabelece infraestruturas básicas que precisam ainda ser implementadas para que possam ser efetivamente utilizadas. Como complicador adicional, observa-se que diferentes estratégias de preservação e diferentes tipos de informação digital exigem tipos distintos de metadados.

#### **4 QUAIS SÃO AS INFORMAÇÕES NECESSÁRIAS PARA A PRESERVAÇÃO DIGITAL?**

A definição dos tipos e dos contornos das informações necessárias para se instruir corretamente os processos de preservação digital foi objeto de grandes discussões num passado recente. Porém, apesar dos inúmeros pontos de tensões, os debates foram capazes de estabelecer um consenso em torno de cinco grandes categorias de informação. Essas categorias são materializadas por uma descrição aprofundada e ampla dos aspectos técnicos, custodiais e legais dos recursos digitais que devem ser traduzidos por metadados de preservação. Resumidamente, são as seguintes: 1) *proveniência* – os metadados de preservação devem registrar informações sobre a história do objeto desde sua origem, traçando a sua cadeia de custódia e de propriedade; 2) *autenticidade* – os metadados de preservação devem incluir informações suficientes para validar que o objeto é

de fato o que diz ser e que não sofreu alterações – intencionais ou não - não documentadas; 3) *atividades de preservação* – os metadados de preservação devem documentar as ações tomadas ao longo do tempo para preservar o objeto digital e as consequências dessas ações sobre aparência, usabilidade e funcionalidades do objeto; 4) *ambiente técnico* – os metadados de preservação devem descrever as dependências técnicas necessárias para a apresentação e uso dos objetos digitais, tais como *hardware*, sistema operacional e *software* de aplicação; 5) *gestão de direitos* – os metadados de preservação devem registrar todos os itens relacionados às questões de propriedade intelectual que limitem as ações de preservação, de disseminação e uso por parte de usuários de hoje e do futuro (LAVOIE; GARTNER, 2005).

Quando pensamos na estruturação das informações necessárias para preservação digital na forma de esquemas de metadados, muitos fatores devem ser levados em consideração. Entretanto, três deles são particularmente importantes, consideradas as idiossincrasias da área: *abrangência* – o esquema deve ter uma amplitude tal, em termos de escopo e de profundidade, que considere as necessidades presentes e futuras de preservação do sistema de repositório considerado; *orientação para a implementação* – o esquema deve ser projetado tendo como perspectiva os níveis práticos de implementação e a possibilidade de adaptação a sistemas automatizados voltados para gerir e assinalar metadados; *interoperáveis* – os esquemas devem ser pensados para promover e facilitar as transações entre diversos fatores que envolvam o objeto digital e os seus diversos metadados ao longo do seu ciclo de vida, por exemplo, submissão a um repositório, disseminação para um usuário ou transferência para outro repositório (LAVOIE; GARTNER, 2005).

Mas a aplicação de esquemas de metadados de preservação é uma aposta que tem como referência um cenário postulado para o futuro e torna-se um espaço amplo para incertezas e conjecturas.

Um dos principais desafios no desenvolvimento de esquemas de metadados de preservação é antecipar que informação será realmente necessária para assistir uma atividade específica de preservação digital. A extensão e a profundidade das que são exigidas para apoiar uma determinada atividade de preservação digital é função direta de algumas variáveis importantes, por exemplo: a intensidade de preservação aplicada a um dado objeto digital arquivado, ou seja, o número de características que devem ser preservadas – funcionalidades, usabilidade, aparência,

autenticidade, etc.; a duração do arquivamento; a complexidade do objeto digital; ou mesmo a base de conhecimento da comunidade a quem a informação se dirige. A decisão sobre como será aplicado o esquema define a política de preservação de um dado repositório. Por exemplo, um repositório de teses e dissertações, cujos materiais arquivados são caracteristicamente textos, tem exigências diferentes de um repositório de objetos multimídias.

Uma vez que um esquema de metadados de preservação é desenvolvido e implementado, fica difícil julgar sua efetividade *a priori*, já que uma avaliação só poderá ser realizada no futuro. Ao contrário dos metadados voltados para apoiar a descoberta de recursos, que podem ser prontamente testados e refinados para que melhorem as métricas de relevância e precisão dos resultados de busca, a adequação de um conjunto de elementos de metadados de preservação só pode ser determinada muito tempo depois da sua implementação. Só nesse momento, se pode avaliar se as informações foram excessivas ou – o que pode ser desastroso – insuficientes para garantir a preservação de longo prazo.

## **6. O MODELO DE REFERÊNCIA OAIS – OPEN ARCHIVAL INFORMATION SYSTEM**

No movimento entre teoria e prática nos espaços da preservação digital, dois pontos extremos são referenciais e significativos para o desenvolvimento de uma infraestrutura voltada para a implementação de metadados de preservação: no extremo conceitual está o OAIS Information Model e no prático, o PREMIS Data Dictionary; entre eles há um campo vasto onde várias iniciativas importantes se sucedem e se sobrepõem. Nessa seção trataremos do modelo OAIS, na seguinte, do PREMIS.

O modelo de referência OAIS é uma infraestrutura conceitual que descreve o ambiente, as interfaces externas, os componentes funcionais e os objetos de informação, associados com um sistema responsável pela preservação de longo prazo de materiais digitais. O modelo é uma tentativa de oferecer uma infraestrutura comum que pode ser usada para se compreender melhor os desafios que os repositórios precisam enfrentar; define também uma linguagem comum de alto

nível que serve de instrumento para facilitar a discussão entre as diferentes comunidades interessadas no problema de preservação digital (DAY, 2004; SARAMAGO, 2004).

O OAIS foi aprovado como uma norma internacional em 2003<sup>13</sup>, porém, antes disso, ele já era amplamente adotado por comunidades importantes na área de preservação digital que definiam seus repositórios como aderentes ao OAIS. A elaboração do Modelo foi coordenada pelo Consultive Committee for Space Data Systems (CCSDS)<sup>14</sup>, vinculado a NASA<sup>15</sup>, como parte de uma iniciativa da International Organization for Standardization<sup>16</sup> (ISO) para o desenvolvimento de normas capazes de regular a preservação de longo prazo de dados originados por satélites e missões espaciais. Porém, o OAIS foi desenvolvido como um modelo genérico, aplicável a qualquer contexto de preservação digital. Nessa direção, a norma descreve um enquadramento conceitual para um repositório digital genérico, aberto, interoperável e com garantias de confiabilidade (SARAMAGO, 2004), que se autodefine “uma organização de pessoas e sistemas que aceitaram a responsabilidade de preservar a informação e torná-la disponível para uma comunidade-alvo” (CCSDS, 2002, p.1-11, tradução nossa).

Em primeiro plano, o OAIS define duas infraestruturas abstratas: *um modelo funcional* e *um modelo de informação*. O modelo funcional é compreendido como um conjunto de atividades que devem ser desempenhadas por um repositório OAIS, seja ele digital ou não; a infraestrutura funcional especificada no documento inclui admissão, armazenamento, gestão de dados, planejamento da preservação, administração e acesso. O modelo de informação define as informações, expressas por metadados, necessárias para a preservação de longo prazo e acesso aos objetos armazenados num sistema baseado no OAIS. O modelo de informação constitui uma conceitualização dos objetos de informação incorporados, armazenados e disseminados por um repositório digital orientado para a preservação (CCSDS, 2002).

O OAIS define ainda o ambiente onde interagem os protagonistas envolvidos em todo o ciclo: o produtor – papel desempenhado pelas pessoas ou sistemas que fornecem a informação que deve ser preservada; a administração – papel desempenhado por quem estabelece as políticas gerais do repositório; o consumidor (usuário) - papel desempenhado por pessoas ou sistemas que

---

<sup>13</sup> ISO Standard 14721:2000. No Brasil a norma foi traduzida e publicada pela ABNT como ABNT NBR 15.472:2007 - Sistema Aberto de Arquivamento de informações (SAAI)

<sup>14</sup> <<http://public.ccsds.org/default.aspx>>

<sup>15</sup> < <http://www.nasa.gov/home/index.html> >

<sup>16</sup> <<http://www.iso.org/>>

interagem com os serviços do repositório com o propósito de identificar e adquirir a informação preservada que deseja; uma classe especial de consumidores, chamada de *comunidade-alvo*, é definida como o conjunto de consumidores que devem ser capazes de compreender a informação preservada. (CCSDS, 2002, p.2-2).

Inicialmente, pode haver um estranhamento em relação aos termos adotados pelo modelo, mas no documento OAIS há uma intencionalidade óbvia em se adotar um discurso independente de áreas específicas. Isso consubstancia a ideia de um modelo genérico e de domínio amplo de aplicação que torne possível a participação de instituições não arquivísticas nos processos de preservação digital.

Não obstante a amplitude da norma que toca em vários aspectos relevantes, a questão de metadados definida no seu escopo é determinante para a área de preservação digital. “O OAIS vem exercendo uma profunda influência no desenvolvimento da arte e da ciência da preservação digital e na área de metadados de preservação é onde este impacto é especialmente evidente”, ratificam LAVOIE e GARTNER (2005, p.9, tradução nossa).

Quando oferece uma descrição de alto-nível dos tipos de informação que fluem no espaço onde se desenrolam os processos do que chamamos de preservação digital, o OAIS torna evidente o vínculo entre metadados e preservação digital e, dessa forma, reconstrói a ideia de metadados de preservação em bases mais sólidas. Como desdobramento, o modelo de informação OAIS vem constituindo o fundamento comum para a orientação e o desenvolvimento da maioria das iniciativas de metadados de preservação surgidas nos últimos anos.

## **6.1 O Modelo de Informação OAIS**

O modelo de informação definido no escopo do documento OAIS especifica o espectro de diferentes tipos de informação – ou metadados - exigidos para assegurar a preservação por um período indefinido de tempo, que pressupõe ainda o acesso aos conteúdos e a sua correta interpretação pelas comunidades interessadas. Os tipos de metadados, que são necessários para a preservação, são definidos como parte de uma Taxonomia de Classes de Objetos de Informação (CCDS, 2002, p. 4-23).

O pressuposto básico do Modelo de Referência OAIS é que um recurso de informação tenha dois componentes: o objeto que precisa ser preservado e as informações que tornem o objeto compreensível para os usuários do repositório OAIS; mais formalmente, significa dizer que todo *Objeto de Informação* é composto por *Objetos de Dados* – que pode ser um objeto físico (por exemplo, uma amostra lunar) ou um objeto digital (seqüências de *bits*), e por *Informação de Representação*, que permite a completa interpretação dos dados em informações com significado (CCSDS, 2002, p.4-19).

Para um objeto digital, que é composto por uma ou mais seqüências de *bits*, o propósito da Informação de Representação é converter seus *bits* em conteúdos mais expressivos aos sentidos, ou seja, em texto, em imagem, em tabela, etc. Isso é realizado através da descrição de formatos de arquivo ou de conceitos de estruturas de dados aplicado à seqüência de *bits*. Pode incluir também informações adicionais necessárias para estabelecer significados particulares de um conteúdo (DAY, 2005).

Este dispositivo de reconstituição do significado da informação assume dois tipos: *informação estrutural* e *informação semântica*.

A informação estrutural inclui especificações, tais como formato dos dados, descrição do ambiente de *hardware* e de *software* em que os dados foram criados; já a informação semântica acrescenta significado à estrutura de dados identificada através da informação estrutural. Por exemplo, a informação estrutural identifica que a seqüência de bits é um texto ASCII, enquanto a informação semântica indica que o texto se encontra escrito em língua inglesa (SARAMAGO, 2004).

A ideia de Objeto de Informação composta por Objeto de Dados e Informação de Representação é aplicada a todo o tipo de informação discutida no âmbito do OAIS. Isso implica na necessidade de definir estruturas lógicas que vinculem o conteúdo a ser preservado à diversidade de metadados que apoiarão a gestão da sua preservação. Decorre daí a ideia de *pacote de informação*.

No ambiente de um repositório aderente à norma OAIS, os fluxos de informação se realizam por meio de unidades discretas chamadas *Pacotes de Informação* - contêineres que encapsulam logicamente os conteúdos, objeto da preservação e os metadados associados a eles (CCSDS,

2002, p2-5). Esse é um conceito-chave subjacente a todos os processos que se desenrolam no âmbito do modelo OAIS.

A norma define três tipos de pacotes de informação: *pacote de informação de submissão*<sup>17</sup>, formado pelo conteúdo e metadados que são submetidos pela entidade externa, *Produtor*, ao repositório no momento do depósito; *pacote de informação de armazenamento*<sup>18</sup>, formado pelo conteúdo e pelos metadados que são efetivamente armazenados e gerenciados pelo repositório por longo prazo; o *pacote de informação de disseminação*<sup>19</sup>, que é o conteúdo e os metadados entregues pelo repositório em resposta a uma requisição de acesso demandada pelo usuário, ou melhor, pelo *Consumidor*.

Deve ficar claro que o pacote de informação de armazenamento é o pacote destinado à preservação de longo prazo; ele é um contêiner que agrega quatro tipos de objetos de informação que circunscrevem os vários tipos de informações necessárias para a preservação de longo prazo (CCSDS, 2002, p.2-6), ou seja:

- *Informação de conteúdo* – é a informação que o repositório tem obrigação de preservar, inclui a *informação de representação*, que são informações necessárias à apresentação e à interpretação da cadeia de bits que constituem o objeto armazenado como informação com significado para uma determinada comunidade alvo;
- *Informação de descrição de preservação* – informação que apoia e documenta a preservação dos objetos arquivados no repositório;
- *Informação de empacotamento* – informação que agrega todos os componentes de um pacote de informação – conteúdo e seus metadados - numa única unidade lógica;
- *Informação descritiva* – informação que apoia o usuário na descoberta e na recuperação de objetos armazenados no repositório.

A *informação de descrição de preservação*, identificada pelo OAIS pela sigla PDI<sup>20</sup>, é o tipo de informação que nos interessa nesse momento. O PDI está “especificamente focado na descrição do estado, tanto passado quanto presente, da Informação de Conteúdo, assegurando que ela está

---

<sup>17</sup> Do inglês *Submission Information Package* (SIP).

<sup>18</sup> Do inglês *Archival Information Package* (AIP).

<sup>19</sup> Do inglês *Dissemination Information Package* (DIP)

<sup>20</sup> Sigla para *Preservation Description Information*

univocamente identificada e que não sofreu alterações não documentadas” (CCSDS, 2002, p.4-27, tradução nossa). A Taxonomia de Classes de Objeto de Informação do OAIIS detalha a *informação de descrição de preservação* em quatro grupos distintos de dados, definidos como se segue (CCSDS, 2002, p.4-28):

- *Informação de referência* – tem origem na necessidade de identificar e de localizar um objeto ao longo do tempo para manter a sua integridade; a referência identifica ou, se necessário, descreve um ou mais mecanismos usados para assinalar identificadores aos objetos armazenados, de forma que eles possam ser identificados inequivocamente interna e externamente ao repositório. Por exemplo, um identificador local (um número de chamada) e um DOI<sup>21</sup> ou um ISBN<sup>22</sup>; pode incluir ainda informações que descrevem o objeto, por exemplo, um resumo.
- *Informações de contexto* – está relacionado ao fato de que muitos objetos não podem ser adequadamente interpretados sem a compreensão do seu contexto; informação que documenta o relacionamento do objeto armazenado e seu ambiente; isso inclui a motivação da criação do objeto e como ela se relaciona com outros conteúdos; circunscreve as dependências técnicas – hardware, software, *linkage*, etc. - inclui ainda modo de distribuição, por exemplo, via rede.
- *Informação de proveniência* – refere-se ao princípio de que parte da integridade de um objeto depende da sua história; informação que documenta a história do objeto armazenado; pode incluir informações sobre sua fonte ou origem, sua cadeia de custódia; registra também as ações de preservação sofridas pelo objeto e seus efeitos, por exemplo: as migrações efetuadas.
- *Informação de fixidade* – refere-se a qualquer informação que documenta mecanismos particulares de autenticação usados para assegurar que o objeto armazenado não sofreu nenhuma alteração não documentada, e que sua integridade não foi comprometida, por exemplo, assinaturas digitais e *checksums*.

---

<sup>21</sup> Digital Object Identifier - <<http://www.doi.org/>>

<sup>22</sup> International Standard Book Number - <[http://www.bn.br/portal/?nu\\_pagina=26](http://www.bn.br/portal/?nu_pagina=26)>

Esses grupos de informação – que formam as bases das principais estruturas de metadados de preservação - são baseados em categorias definidas pelas discussões apresentadas em 1996 no relatório Task Force on Archiving of Digital Information comissionados pela Commission on Preservation and Access (CPA) e pela Research Library Group<sup>23</sup> (RLG), que registra textualmente que “no ambiente digital, as características que determinam a integridade da informação e merecem uma atenção especial para propósitos de arquivamento incluem: conteúdo, fixidade, referência, proveniência e contexto” (CPA/RLG, 1996, p.12, tradução nossa). Os tipos de informações explicitadas pela taxonomia presente no modelo de informação OAIS podem ser interpretados como a descrição mais geral de metadados necessários para instruir a preservação de longo prazo e o uso de materiais digitais. Essas informações estabelecem um ponto de partida para a maioria dos esforços subsequentes em desenvolver esquemas formais de metadados.

## **6.2 Aplicações do Modelo de Informação OAIS**

Enquanto um modelo de referência, o OAIS não toca nos níveis de implementação (CCSDS, 2002, p.1-3); cada comunidade interessada deve aplicar o modelo – incluindo o modelo de informação – no seu contexto técnico e organizacional, adequando-o as suas especificidades e objetivos. Ainda no seu papel de uma descrição de alto nível, a norma não transmite pressupostos sobre os tipos de recursos digitais manuseados pelo repositório e nem acerca das especificações tecnológicas adotadas por ele para cumprir os seus objetivos de preservação e acesso de longo prazo (SARAMAGO, 2004).

Entretanto, a demanda por desenvolvimento de soluções operacionais está refletida na longa lista de instituições envolvidas na criação de conjuntos de elementos de metadados para apoiar a preservação digital. Michael Day (2003) nos informa que a maioria dessas implementações surge em três contextos distintos que, porém, possuem o interesse comum pela preservação digital: bibliotecas nacionais e de pesquisa, projetos de digitalização e arquivos. Algumas dessas

---

<sup>23</sup> <<http://www.oclc.org/programs/about/default.htm>>

iniciativas, originadas por instituições internacionais de maior renome, são projetos com desdobramentos significativos para a área. Vamos nos ater, neste momento, às iniciativas que têm em comum uma fundamentação – embora em graus variados – inspirada no modelo OAIS.

Uma das primeiras respostas práticas ao desafio foi dada pelo National Library of Australia (NLA)<sup>24</sup>, tendo como ambiente o repositório de publicações eletrônicas PANDORA<sup>25</sup>, sigla para Preserving and Accessing Networked Documentary Resources of Australia; logo após, a minuta de outro conjunto de elementos foi publicada no Reino Unido, no âmbito do projeto CEDARS<sup>26</sup> (CURL Exemplars in Digital Archives); o projeto NEDLIB<sup>27</sup> (Networked European Deposit Library) desenvolveu um sistema de depósito para bibliotecas eletrônicas e tentou definir, nesse contexto, um conjunto mínimo de metadados que seria necessário para apoiar a gestão da preservação.

LAVOIE e GARTNER (2005) observam que esses primeiros esforços resumem uma tendência por natureza altamente especulativa, dado que procuravam antecipar os elementos de metadados de preservação necessários para sustentar as iniciativas programáticas de preservação digital que iriam emergir no futuro. Não havia consenso sobre questões básicas, tais como que tipos de informações seriam necessárias e como elas poderiam ser usadas para apoiar os processos de preservação digital. Por outro lado, os projetos importantes que se seguiram – por exemplo, os conjuntos de elementos produzidos pela OCLC<sup>28</sup>, pela National Library of New Zealand<sup>29</sup> e pela University of Edinburg<sup>30</sup> - estavam mais proximamente alinhados com o planejamento e a implementação de sistemas de repositórios digitais e se beneficiaram amplamente da fundamentação estabelecida pelos primeiros conjuntos de elementos.

Na trajetória que se delineava, tornava-se imperativo, para a área de preservação digital, harmonizar os três esquemas referenciados acima – NLA, CEDARS e NEDLIB - em uma infraestrutura única. Nessa direção, por volta do ano 2000, a OCLC e o Research Library Group (RLG) convocaram um grupo de trabalho internacional, que reunia expertise de vários domínios e organizações, para endereçar os novos desenvolvimentos na área. O grupo produziu dois

---

<sup>24</sup> <<http://www.nla.gov.au/>>

<sup>25</sup> <<http://pandora.nla.gov.au/>>

<sup>26</sup> <<http://www.rluk.ac.uk/projects>>

<sup>27</sup> <<http://nedlib.kb.nl/>>

<sup>28</sup> <<http://www.dpconline.org/docs/reports/dpctw05-01.pdf>>

<sup>29</sup> <<http://www.natlib.govt.nz/catalogues/library-documents/preservation-metadata-revised>>

<sup>30</sup> <<http://www.lib.ed.ac.uk/sites/digpres/metadataschema.shtml>>

relatórios que constituíram documentos determinantes para o avanço na direção de uma efetiva implementação fundamentada no OAIS, são eles: “Preservation metadata for digital object: a review of the state of the art” (OCLC/RLG, 2001) e o “Preservation metadata and the OAIS Information Model: a metadata framework to support the preservation of digital object” (OCLC-RLG, 2002).

O primeiro documento – um livro branco - sintetizava o estado da arte em metadados de preservação digital, oferecia uma definição para eles, descrevia os papéis dessa classe de metadados no processo de preservação, ao mesmo tempo em que revisava as iniciativas existentes, identificando convergências e divergências. A tarefa seguinte estava fundamentada sobre os alicerces consolidados por este livro branco e tinha como resultado o desenvolvimento de uma infraestrutura de metadados de preservação abrangente e de larga aplicação baseada nas categorias de informação especificadas no modelo de informação do OAIS.

“A infraestrutura produzida pelo grupo de trabalho efetivamente substituiu o conjunto de elementos desenvolvido pelas iniciativas anteriores e representaram um ponto de partida importante para a futura implementação prática de metadados de preservação” (DAY, 2003, p.5, tradução nossa), ao mesmo tempo em que deixava óbvio que a colaboração e o consenso formam a pedra de toque para superar os desafios e as incertezas da preservação digital. Entretanto, ainda era necessário um esforço considerável antes que fosse possível implementar operacionalmente esquemas de metadados de preservação para repositórios particulares, posto que algumas questões importantes sobre os metadados e seus usos ainda precisavam ser respondidas. Por exemplo: de todas as informações cobertas pela infraestrutura, qual é o subconjunto de informações essenciais para preservação de longo prazo? Como essas informações podem ser traduzidas em elementos implementáveis de metadados de preservação? Como os metadados de preservação podem ser criados e mantidos no âmbito operacional de um sistema de arquivamento digital? (LAVOIE; GARTNER, 2005).

Para responder questões como essas, a OCLC e a RLG patrocinaram, logo em seguida, um novo grupo de trabalho chamado PREMIS – sigla para Preservation Metadata: Implementation Strategies<sup>31</sup> - com o objetivo de detalhar os aspectos práticos de implementação dos metadados

---

<sup>31</sup> <<http://www.loc.gov/standards/premis/>>

de preservação no contexto de sistemas de preservação digital. É sobre isso que discutiremos brevemente a seguir.

## **7 PREMIS: O MODELO OAIS EM AÇÃO**

O objetivo subjacente à idéia de constituir o Grupo de Trabalho PREMIS era delinear uma ferramenta concreta, uma ponte, que pudesse superar o abismo entre a teoria e a prática na área de metadados de preservação digital; o que também pode ser traduzido por colocar em ação os conceitos preconizados pela infraestrutura de alto nível fixada pelo Modelo de Informação do OAIS. Nessa direção, o PREMIS se estabeleceu tendo como base o consenso extraído das experiências acumuladas de muitas e variadas instituições – museus, bibliotecas, arquivos, governo e iniciativa privada – e a *expertise* dos principais profissionais da área, provenientes da Austrália, Nova Zelândia, Estados Unidos, Grã-Bretanha, Holanda e Alemanha. O empreendimento foi inicialmente planejado para um ano, porém se desdobrou por mais outro. Os resultados, entretanto, compensaram o alongamento dos prazos: o Grupo de Trabalho desenvolveu um conjunto de elementos de metadados altamente refinados, que potencialmente servia de fundamento para possíveis implementações (McCULUN, 2005).

O esforço considerava vários objetivos relacionados. Porém, o interesse do Grupo de Trabalho convergia de forma contundente para dois pontos que sintetizavam o que se esperava do OAIS, enquanto uma fundamentação para a prática da preservação digital, para o intercâmbio de informações de preservação e para a interoperabilidade entre repositórios. Esses pontos eram os seguintes:

- Tomando como ponto de partida a infraestrutura delineada anteriormente, definir um conjunto essencial de elementos de metadados de preservação que seja implementável e de larga aplicação; esse núcleo essencial de metadados deve ser apoiado por um dicionário de dados, que será desenvolvido para oferecer diretrizes e recomendações para o preenchimento e para a gestão dos elementos de metadados.

- Identificar e avaliar estratégias alternativas para codificar, armazenar, gerenciar e intercambiar metadados de preservação, especialmente os essenciais, no contexto de um sistema de repositório digital.

O trabalho do Grupo começou pelo levantamento dos projetos de repositórios digitais em operação e ainda os planejados, tendo como objetivo identificar as práticas correntes e as tendências para projetos digitais. Dentre os vários aspectos endereçados – missão, comunidade de usuários, serviços, fundos de financiamento, gestão de direitos e conteúdos – estavam, naturalmente, interrogações sobre como os metadados estavam sendo usados para apoiar os processos, as funções e as políticas do repositório. (LAVOIE; GARTNER, 2005). O levantamento obteve 48 respostas, originadas principalmente por bibliotecas, arquivos e museus provenientes de 13 países diferentes. Os resultados do *survey* foram sumarizados no relatório “Implementing preservation repositories for digital materials: current practice and emerging trends in the cultural heritage (OCLC/RLG, 2004).

Contudo, a principal materialização do trabalho do PREMIS foi o relatório de 237 páginas lançado em maio de 2005, intitulado “Data dictionary for preservation metadata: final report of the PREMIS Work Group (OCLC/RLG, 2005).

## **7.1 Dicionário de dados PREMIS**

O coração e a alma deste relatório é o PREMIS Data Dictionary, traduzido aqui por Dicionário de Dados PREMIS. Trata-se de um guia abrangente que define um conjunto de metadados necessários para apoiar a preservação digital de longo prazo.

O Dicionário de Dados não tem como objetivo definir todos os elementos possíveis de metadados de preservação, verdadeiramente ele se concentra no núcleo básico de elementos que a maioria dos repositórios precisa compreender para apoiar a preservação de longo prazo; esse núcleo é chamado de *metadados essenciais*.

O relatório inclui complementarmente vários outros textos e ferramentas: os “tópicos especiais” que discutem aspectos relacionados ao Dicionário de Dados; um glossário; e um conjunto de exemplos que ilustram o uso do Dicionário de Dados para vários materiais em diferentes contextos de preservação digital. O Grupo de Trabalho desenvolveu também um conjunto de esquemas XML<sup>32</sup> para apoiar o uso do Dicionário de Dados por instituições que gerenciam e intercambiam metadados de preservação que estejam em conformidade com a proposta do PREMIS.

Rigorosamente, o Dicionário de Dados não define elementos de metadados e sim unidades semânticas. Essa diferença é sutil, porém importante: uma unidade semântica é uma peça de informação ou de conhecimento, enquanto um elemento de metadados é uma forma definida de representar essa informação em um registro de metadados, em um esquema ou numa base de dados. Nessa direção, o PREMIS não especifica como os metadados devem ser representados em um sistema, ele simplesmente define o que o sistema precisa entender e o que ele deve ser capaz de exportar para outros sistemas (CAPLAN, 2009).

O Dicionário de Dados está organizado em torno de um modelo de dados (figura 1) que relaciona cinco entidades que têm papéis associadas com a preservação digital, são elas: *Entidade Intelectual*, *Objeto*, *Evento*, *Agente* e *Direitos*. O PREMIS as define da seguinte forma:

- *Entidade intelectual* – um conjunto coerente de conteúdos que é reconhecido como uma unidade, por exemplo, livros, artigos, bases de dados;
- *Objeto* – uma unidade discreta de informação em forma digital, constituindo o que realmente é armazenado e gerenciado pelo repositório, por exemplo, um arquivo PDF. As unidades semânticas para Objetos podem ser especificadas em três níveis: cadeia de bits (*bitstream*), arquivos (*files*) e o conjunto de arquivos que completam a apresentação de uma Entidade Intelectual, ou seja, a representação (*representation*).
- *Evento* – são ações que envolvem ou afetam os objetos no repositório, por exemplo, uma ação de migração;
- *Agente* – é uma pessoa, organização ou programa de computador que desempenha papéis associado com um Evento ou declarações de Direitos;

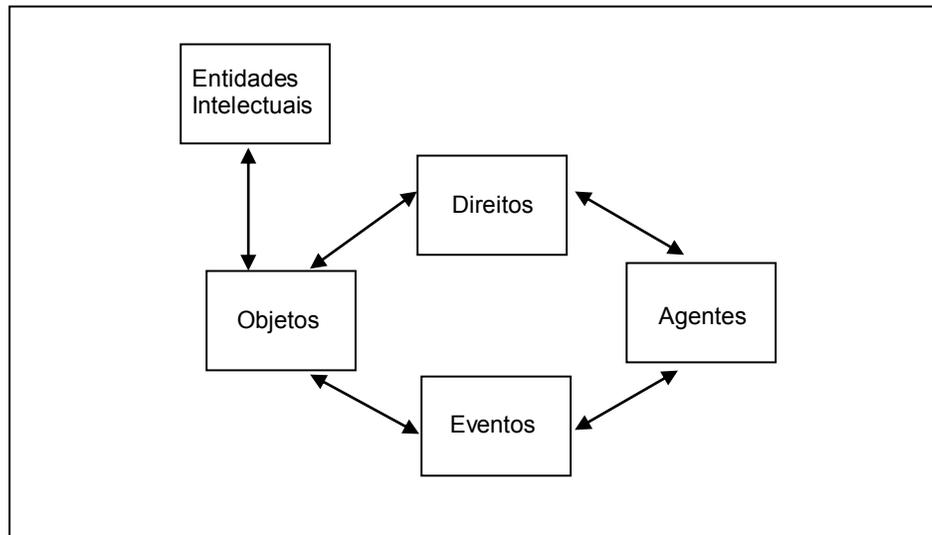
---

<sup>32</sup> <<http://www.loc.gov/standards/premis/schemas.html>>

- *Direitos* – são direitos e permissões vinculadas ao Objeto relevantes para a preservação, por exemplo, permissão para se fazer uma cópia em PDF.

O Dicionário de Dados oferece uma descrição detalhada dos metadados associados com cada uma das entidades, entretanto os metadados para Entidades Intelectuais são considerados fora do escopo do PREMIS, dado que a estas informações já são supridas pelos esquemas focados em metadados descritivos (CAPLAN, 2009; LAVOIE; GARTNER, 2005). Por exemplo, MARC para materiais bibliográficos.

Intencionalmente, o Grupo de Trabalho PREMIS não tratou de alguns aspectos bem conhecidos da preservação digital, tal como o detalhamento dos metadados técnicos para diferentes mídias; somente os metadados técnicos que são geralmente aplicados transversalmente a formatos de arquivos foram trabalhados pelo Grupo. Outra importante consideração adotada pelo PREMIS é que os metadados especificados devem ser, tanto quanto possível, assinalados e usados automaticamente. Isso leva preferencialmente para a escolha de valores extraídos de listas contendo formas padronizadas, ao invés de descrição textual (MCCALLUM, 2005).



**Figura 1:** Modelo de Dados do PREMIS

Lavoie e Gartner (2005, p.14) observam que “há ainda muito trabalho a ser feito, especialmente em termos de testar o Dicionário de Dados em diferentes domínios e contextos de preservação digital”; eles concluem refletindo que no futuro, a ampla adoção do Dicionário de Dados pode ajudar no estabelecimento de práticas padronizadas voltadas para a gestão de metadados de Enc. Bibli. R. Eletr. Bibliotecon. Ci. Inf., ISSN 1518-2924, Florianópolis, v. 15, n. 30, p.1-31, 2010.

preservação que enfatizem a interoperabilidade de repositórios digitais distribuídos em redes. A adoção de padrões pode ainda gerar uma economia potencial possibilitada pela prática de compartilhar e reusar determinadas formas de metadados de preservação entre repositórios digitais.

Nessa direção, o PREMIS Maintenance Activity desenvolveu um esquema XML que corresponde diretamente ao Dicionário de Dados, viabilizando que o PREMIS seja usado para intercâmbio de metadados representado em XML.

## **8 METS: EMPACOTANDO OS METADADOS DE UM OBJETO DIGITAL**

Não há dúvidas de que um objeto digital vai acumulando uma quantidade crescente de metadados de todo o tipo ao longo de tempo; somadas aos metadados de preservação muitas outras formas são incorporadas, tais como metadados para a descoberta de recursos e administrativos. Este fato nos coloca diante de uma questão crítica: como todos esses metadados podem estar organizados e vinculados ao objeto correspondente? Algumas soluções foram propostas na forma de infraestruturas para empacotamento de metadados, dentre elas estão o MPEG-21<sup>33</sup> e o METS, sigla para Metadata Encoding Transmission Standard. No contexto que nos interessa, o mais importante é a norma METS, posto que ela foi projetada por iniciativa da Digital Library Federation (DLF) para implementar os pacotes de informação referenciados pelo Modelo de Referência OAIS (LAVOIE, 2004).

O METS é um esquema XML que oferece um mecanismo flexível para codificar todos os tipos de metadados associados a um objeto digital – descritivos, administrativos, estruturais - e para exprimir as ligações complexas entre esses metadados no ambiente de um repositório. Por conseguinte, o METS estabelece um padrão útil para a gestão de objetos digitais no âmbito de um repositório e o intercâmbio deles entre repositórios (ou entre repositórios e seus usuários); além do mais, oferece a possibilidade de associar um objeto digital com comportamentos ou serviços. Dessa forma, um documento METS pode ser usado para estruturar Pacotes de Informação de

---

<sup>33</sup> <<http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm>>

Submissão, Pacotes de Informação de Arquivamento e Pacotes de Informação de Disseminação, que é a forma como as informações são gerenciadas e fluem no contexto do Modelo de Referência OAIS (LIBRARY OF CONGRESS, 2009).

Um documento METS compreende cinco principais seções:

- Grupo de arquivos – é um inventário de todos os arquivos associados com o objeto digital e de suas versões eletrônicas.
- Metadados Administrativos – essa seção aninha as informações técnicas sobre: como os arquivos foram criados e armazenados, a gestão de direitos, o objeto original da qual o objeto deriva e a proveniência dos arquivos que compõem o objeto. Pode apontar para metadados externos ao documento METS.
- Metadados Descritivos – essa seção inclui informações sobre o conteúdo intelectual do item – incluindo informações bibliográficas - necessárias para a sua recuperação e avaliação por parte do usuário. Essa seção pode apontar para metadados externos ao documento METS, por exemplo, um registro MARC num catálogo *on-line* (OPAC).
- Mapa Estrutural – indica de forma hierárquica como os vários componentes do item se relacionam mutuamente, permitindo, dessa forma, que seus elementos constituintes possam ser navegados pelos usuários.
- Comportamento – essa seção pode ser usada para associar comportamentos executáveis com o conteúdo no objeto METS.

De muitas formas, o METS representa uma solução que se enquadra nas exigências de estabilidade da preservação digital. Em primeiro lugar, um documento METS está escrito em XML, que há muito tem sido consensualmente reconhecido por todos os domínios como uma forma robusta e legível para o arquivamento de metadados; depois, enquanto uma linguagem não proprietária, o XML pode assegurar que a informação, por ele codificada, não será dependente de nenhum pacote específico de *software* e, portanto, não sofrerá – ou sofrerá menos - as consequências da obsolescência tecnológica que ameaça as aplicações vinculadas a programas. Portanto, os metadados arquivados em dispositivos XML, tal como o padrão METS, deverão estar prontos para uso pelos mecanismos futuros de disseminação e de intercâmbio com outros repositórios (LAVOIE; GARTNER, 2005).

## 9 À GUIA DE CONCLUSÃO

Os metadados têm um papel de fundamental importância na organização e no acesso às informações nos sistemas tradicionais, como nas coleções de livros de uma biblioteca ou nos ambientes informacionais baseados em redes de computadores, como é a própria *web*. Entretanto, o conceito de metadado pode ser expandido para apoiar a gestão de objetos digitais, cujo escopo inclui os processos de preservação digital de longo prazo.

Progressivamente, essa idéia foi se consolidando. Hoje, há um consenso absoluto de que os conteúdos digitais que precisam ser acessados e compreendidos no futuro devem estar acompanhados de dados e informações, expressos na forma de metadados, que tornem viável a sua acessibilidade, integridade e autenticidade.

Nessa direção, iniciando-se na década de 1990, inúmeros projetos e iniciativas vêm enfrentando o desafio de dimensionar o papel dos metadados no apoio às atividades de preservação digital e de identificar quais são as informações necessárias para tal. Esses esforços têm como característica comum o desenvolvimento baseado no consenso e na cooperação.

A universalidade do problema da fragilidade da informação digital, bem como a convergência de interesses das diversas instituições de patrimônio digital – bibliotecas, museus e arquivos - falam a favor da colaboração e da construção do consenso para resolver os desafios e as incertezas de gerenciar materiais digitais por longo prazo. Numa trajetória evolutiva, diretrizes, padrões, práticas e experiências em implementação estão emergindo e se consolidando baseados em modelos conceituais concebidos num passado recente. O PREMIS, considerada a iniciativa mais importante em metadados de preservação, é uma síntese de tudo isso. Baseado nas experiências acumuladas por muitas instituições, na transversalidade de vários domínios e consolidado pelo consenso, ele representa um passo importante na superação do hiato existente entre a teoria e a prática no domínio da preservação digital.

Por fim, é necessário enfatizar que a relevância e a complexidade do problema da preservação digital podem ser mensuradas pela dependência quase total de dados e de informações digitais de alguns segmentos importantes da sociedade, por exemplo, educação, governo, negócios, pesquisa científica e expressão artística; isso sem falar nas mensagens para o futuro, que são críticas para a

sobrevivência da humanidade, por exemplo, a localização de depósitos de materiais tóxicos. Essa dependência dramática se reflete na urgência pela busca de soluções abrangentes que sejam tecnológicas, econômicas, éticas e legalmente viáveis.

Tudo isso somado transforma a área de pesquisa e da prática em preservação digital um espaço pleno de desafios instigantes para muitos domínios do conhecimento.

### *ANOTHER FACE OF THE METADATA: INFORMATION FOR MANAGEMENT OF THE DIGITAL PRESERVATION*

#### **Abstract**

*The traditional concept of metadata can be expanded to provide a set of information to support the management activities of the preservation of digital materials. This type of metadata, called preservation metadata, is designed to inform and document the process of digital preservation of long-term, assuring that digital content can be accessed and interpreted in the future. In recent years many metadata schemes and infrastructure oriented for digital preservation have been developed; the greatest challenge they face has been to anticipate what information is actually required to support a particular process of digital preservation. The most important and comprehensive initiative in this field is the PREMIS Data Dictionary, developed based on the conceptual infrastructure defined by the OAIS ISO standard. The basic idea of this paper is to review the main concepts, standards and technologies involved in the development of metadata schemes of preservation metadata.*

**Key-words:** *Preservation metadata. Digital preservation. PREMIS. OAIS. METS.*

#### **REFERÊNCIAS**

CAPLAN, Priscilla. **Understanding PREMIS**. Washington D.C.: Library of Congress, 2009. Disponível em: <<http://www.loc.gov/standards/premis/understanding-premis.pdf>>. Acesso em: 30 set. 2009.

CCSDS – Consultative Committee for Space Data System. **Reference Model for an Open Archival Information System (OAIS)**. Blue book (CCSDS 650.0-B-1). Washington, DC : CCSDS, 2002. Disponível em <<http://public.ccsds.org/publications/archive/650x0b1.pdf>>. Acesso em: 30 set. 2009.

CPA/RLG – COMMISSION ON PRESERVATION AND ACCESS; RESEARCH LIBRARIES GROUP. **Preserving digital information: Report of the task force on archiving of digital information**, 1996. 71p. Disponível em: <<http://www.clir.org/pubs/reports/pub63watersgarrett.pdf>>. Acesso em: 11 nov. 2009.

CONARQ – Conselho Nacional de Arquivos. **Carta para a preservação do patrimônio arquivísticos digital:** preservar para garantir o acesso. 2004. Disponível em: <<http://www.conarq.arquivonacional.gov.br/Media/publicacoes/cartapreservpatrimarqdigitalconarq2004.pdf>>. Acesso em 10 dez. 2009.

DAY, Michael. **Preservation metadata.** Bath : UKOLN, University of Bath, 2003. Disponível em: <<http://www.ukoln.ac.uk/metadata/publications/iylim-2003/>>. Acesso em: 11 nov. 2009.

DAY, Michael. **Preservation metadata initiatives: practicality, sustainability, and interoperability.** In: Bischoff, F. M., Ross, S., eds. Metadata in preservation: selected papers ERPANET Seminar at the Archives School Marburg, 2003. Marburg: Archivschule Marburg, 2004, p. 91-117. Disponível em: <<http://opus.bath.ac.uk/14365/1/day-marburg-paper.pdf>>. Acesso em: 01 dez. 2009.

DAY, Michael. **DCC Digital Curation Manual:** installment on “metadata” Bath : University of Bath, 2005. Disponível em <<http://www.dcc.ac.uk/resource/curation-manual/chapters/metadata/metadata.pdf>>. Acesso em: 30 set. 2009.

HOWARD, Besser. Digital longevity. In: SITTS, Maxine (ed.). **Handbook for Digital Projects: A Management Tool for Preservation and Access.** Andover MA: Northeast Document Conservation Center, 2000, p.155-166. Disponível em: <<http://www.gseis.ucla.edu/~howard/Papers/sfs-longevity.html>>. Acessado em: 5 jan. 2010.

LAVOIE, Brian. Implementing metadata in digital preservation systems: the PREMIS activity. **D-Lib Magazine**, v. 10, n. 4, April 2004. Disponível em: <<http://www.dlib.org/dlib/april04/lavoie/04lavoie.html>>. Acesso em: 05 jan. 2010.

LAVOIE, Brian; GARTNER, Richard. **Preservation Metadata.** OCLC, September 2005. Disponível em <<http://www.dpconline.org/docs/reports/dpctw05-01.pdf>>. Acesso em: 30 set. 2009.

LEE, Kyong-Ho *et al.* The state of the art and practice in digital preservation. **Journal of Research of the National Institute of Standards and Technology**, v.107, n.1, p.93-106, Jan.-Feb. 2002. Disponível em: <<http://nvl.nist.gov/pub/nistpubs/jres/107/1/j71lee.pdf>>. Acesso em: 30 out. 2009.

LIBRARY OF CONGRESS. **METS:** an overview & tutorial. 2009. Disponível em: <<http://www.loc.gov/standards/mets/METSOverview.v2.html>>. Acesso em: 12 dez. 2009.

LUKESH, Susan. E-mail and potential loss to future archives and scholarship or the dog that didn't bark. **First Monday**, v. 4, n.9, 1999. Disponível em: <[http://131.193.153.231/www/issues/issue4\\_9/luke/index.html - author](http://131.193.153.231/www/issues/issue4_9/luke/index.html - author)>. Acesso em: 5 jan. 2010.

MARCONDES, Carlos Henrique. Metadados: descrição e recuperação de informação na web. In: MARCONDES, C. H., KURAMOTO, H., TOUTAIN, L. B., SAYÃO, L. F. (org). **Bibliotecas digitais: saberes e práticas.** Salvador: Ed.UFBA, Brasília : IBICT, 2005. p. 97-114.

MCCALLUM, Sally H.. Preservation metadata: what we have and what we need. In: WORLD LIBRARY AND INFORMATION CONGRESS: 71th IFLA General Conference and Council. Oslo, 2005. Libraries - a voyage of discovery. Oslo: IFLA, 2005. p. 1 - 8. Disponível em: <<http://ifla.queenslibrary.org/IV/ifla71/papers/060e-McCallum.pdf>>. Acesso em: 05 jan. 2010.

NISO - National Information Standard Organization. **Understanding Metadata**. Bethesda, MD : NISO Press, 2004. Disponível em: <<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>>. Acesso em: 30 set. 2009.

OCLC/RLG - Online Computer Library Center/ Research Library Group. **Preservation Metadata for digital objects**: a review of the state of the art. 2001. Disponível em: <[http://www.oclc.org/research/activities/past/orprojects/pmwg/presmeta\\_wp.pdf](http://www.oclc.org/research/activities/past/orprojects/pmwg/presmeta_wp.pdf)>. Acessado em: 05 jan. 2010.

\_\_\_\_\_. **Preservation metadata and the OAIS Information Model**: a metadata framework to support the preservation of digital object. 2002. Disponível em: <[http://www.oclc.org/research/activities/past/orprojects/pmwg/pm\\_framework.pdf](http://www.oclc.org/research/activities/past/orprojects/pmwg/pm_framework.pdf)>. Acesso em: 05 jan. 2010.

\_\_\_\_\_. **Data dictionary for preservation metadata**: final report of the PREMIS Working Group. 2005. Disponível em: <<http://www.oclc.org/research/activities/past/orprojects/pmwg/premis-final.pdf>>. Acesso em: 30 set. 2009.

\_\_\_\_\_. **Implementing preservation repositories for digital materials**: current practice and emerging trends in the cultural heritage - A Report by the PREMIS Working Group. September 2004, 66 p. Disponível em: <<http://www.oclc.org/research/activities/past/orprojects/pmwg/surveyreport.pdf>>. Acesso em: 12 dez. 2009.

SARAMAGO, Maria de Lurdes. Metadados para a preservação digital e aplicação do Modelo OAIS. In: CONGRESSO NACIONAL DE BIBLIOTECARIOS, ARQUIVISTAS E DOCUMENTALISTAS, 8., 2004, Estoril. Disponível em: <<http://badinfo.apbad.pt/congresso8/comm2.pdf>>. Acesso em: 30 set. 2009.

SAYÃO, Luis Fernando. Conservação de documentos eletrônicos. In: GRANATO, Marcus; SANTOS, Claudia; ROCHA, Claudia. **Conservação de acervos**. Rio de Janeiro : MAST, 2007, p. 181-204.

*Originais recebidos em: 08/12/2009*

*Aceito para publicação em: 21/06/2010*