

INTRODUÇÃO AO METS – Preservação e Intercâmbio de Objetos Digitais

INTRODUCTION TO METS – Preservation and Interchange of Digital Objects

Nelson de Almeida Rodrigues - nelsonr@unb.br
Mestrando em Ciência da Informação - Universidade de Brasília

Resumo

O Metadata Encoding & Transmission Standard (METS) é um XML Schema elaborado sob o patrocínio da Digital Library Federation e mantido pela Biblioteca do Congresso Norte Americano. Ele permite a criação de uma rica estrutura capaz de registrar, não apenas os múltiplos tipos de metadados usados para descrever o acervo de uma biblioteca digital, como também, os próprios objetos digitais, seja para gerenciá-los, seja para permitir o intercâmbio entre diferentes instituições. Este artigo apresenta os problemas motivadores da sua criação e sua origem histórica no projeto Making of America II. Descreve a finalidade de cada um dos elementos XML que formam o vocabulário desse padrão e ilustra o emprego de alguns deles com exemplos codificados em XML, o que requer um conhecimento básico desta linguagem para melhor compreensão. Expressa sua relação com o modelo Open Archival Information System como um significativo exemplo de sua aplicação e, por fim, aborda outros padrões criados com propósitos semelhantes.

Palavras-chave: Metadados. Biblioteca digital. Preservação digital. METS. Metadata Encoding & Transmission Standard.

1 INTRODUÇÃO

Preservar e facilitar o acesso ao acervo são duas questões fundamentais para as bibliotecas. A história registra a adoção de diferentes abordagens em resposta a esses problemas. Com o advento da Internet, surgiu uma abordagem de solução que tem ganhado força nos últimos anos. Trata-se da digitalização do acervo e de sua consulta através de bibliotecas digitais.

A conversão de obras do acervo físico de uma biblioteca em objetos digitais¹ traz novas necessidades de organização. Considere-se, como exemplo, a digitalização de um livro na qual para cada página sejam gerados um arquivo de imagem e um arquivo com o texto obtido por OCR² da imagem (para permitir buscas textuais). Cada um desses arquivos pode ser visto como um *objeto digital simples* (formado por um único arquivo), entretanto, quando esses objetos são estruturados e agrupados, obtém-se um outro objeto: um livro digital. Este novo objeto é denominado de *objeto digital complexo*, por ser formado pela composição de múltiplos objetos digitais simples. O objeto digital complexo pode ser estruturado de modo a

1 O termo “objeto digital” tem sentido amplo, podendo significar tanto “objeto digital simples” (formado por um único arquivo), como “objeto digital complexo” (formado por múltiplos arquivos).

2 Acrônimo do Inglês *Optical Character Recognition* ou Reconhecimento Ótico de Caracteres, em Português.

permitir consultas on-line às páginas do livro, na forma seqüencial, por capítulo, ou através de buscas textuais. O metadado que antes estava associado a um único elemento (o livro de papel) agora deve ser associado a centenas de elementos (os arquivos de imagem e texto).

A apresentação de obras através de bibliotecas digitais facilita o acesso ao acervo. Cunha (1997) apresenta vasta bibliografia sobre o tema “biblioteca digital” e conclui que desde 1994 esse assunto teve um acentuado crescimento. No entanto, as múltiplas iniciativas de criação de bibliotecas digitais, principalmente a partir da década de 90, cada qual com uma maneira própria de representar seus objetos digitais, trouxe dificuldades para que as bibliotecas pudessem realizar o intercâmbio de elementos de seus acervos digitais, tais como livros e periódicos. Essas iniciativas deram uma nova dimensão aos problemas de migração para o mundo digital e despertaram na comunidade a necessidade de estabelecer padrões. A falta de padrões, segundo Gartner (2002), inibe o crescimento das coleções digitais.

Para viabilizar um crescimento ordenado das coleções digitais, foi elaborado um padrão que permite criar uma rica estrutura capaz de registrar, não apenas os múltiplos tipos de metadados usados para descrever o acervo de uma biblioteca digital, como também, os próprios objetos digitais, seja para gerenciá-los, seja para permitir o intercâmbio entre diferentes instituições. Trata-se do Metadata Encoding & Transmission Standard³ (METS), que é um XML Schema⁴ elaborado sob o patrocínio da Digital Library Federation e mantido pela Biblioteca do Congresso Norte Americano.

Este artigo apresenta os problemas motivadores da criação do METS e sua origem histórica no projeto Making of America II. Descreve a finalidade de cada um dos elementos XML que formam seu vocabulário e ilustra o emprego de alguns deles com exemplos codificados em XML, o que requer um conhecimento básico desta linguagem para melhor compreensão. Expressa sua relação com o modelo Open Archival Information System como um significativo exemplo de sua aplicação e, por fim, aborda outros padrões criados com propósitos semelhantes.

2 HISTÓRICO

Lideradas pela Universidade da Califórnia em Berkeley, a Biblioteca Pública de Nova Iorque e as Universidades de Cornell, do Estado da Pensilvânia e de Stanford se uniram em 1997, apoiadas pela *Digital Library Federation*, para realizar um projeto⁵ que foi denominado de *Making of America II*. De acordo com a *UC Berkeley Library* (1997a), esse projeto se propunha a realizar quatro grandes ações. Primeiro, investigar, refinar e recomendar as melhores práticas para o processo de digitalização. Segundo, definir os metadados a serem utilizados para a navegação, descoberta e exibição de objetos digitais. Terceiro, investigar questões de arquitetura de sistemas para permitir acesso integrado a repositórios distribuídos de objetos digitais. E, por fim, disseminar seus resultados para a comunidade quando de seu término.

O projeto previa a montagem de um ambiente de testes, custeado pela *National Endowment for the Humanities*, onde diversos cenários de conversão de coleções originais para o formato

3 Em português: Padrão para Codificação e Transmissão de Metadados (tradução nossa)

4 Para esclarecimentos sobre XML e as tecnologias a ele associadas, vide <http://www.w3.org/XML/>

5 Há um breve histórico desse projeto em <http://www.diglib.org/standards/dlfmoaii.htm>

digital pudessem ser avaliados e refinados. Para tanto, planejava digitalizar um mínimo de 25.000 páginas de coleções sobre o transporte nos Estados Unidos referentes ao período de 1869 a 1900 (UC BERKELEY LIBRARY, 1997b). A avaliação desses cenários de conversão possibilitaria identificar e padronizar técnicas e procedimentos que tanto reduzissem os custos de montagem de bibliotecas digitais, como permitissem haver interoperabilidade entre elas.

Um padrão de interoperabilidade entre bibliotecas digitais tornaria viável que elas fossem interligadas para formar uma grande Biblioteca Digital Nacional. Isso permitiria que um usuário, acessando qualquer uma dessas bibliotecas digitais, realizasse pesquisas nos acervos de todas elas de forma transparente e descentralizada. Dessa forma, seriam eliminadas as dificuldades de acesso físico aos originais devidas à distribuição geográfica das bibliotecas, assim como aquelas impostas pela fragilidade e raridade dos originais e que impedem um livre manuseio dessas obras. As atividades de pesquisadores e de professores/alunos no processo de ensino e aprendizagem também seriam facilitadas.

No seu término, o projeto *Making of America II* apresentou, dentre outros resultados, um padrão de codificação de metadados para avaliação pela comunidade (HURLEY, 1999). A DLF (2000) relata que, com o passar do tempo, os usuários desse padrão apresentaram a necessidade de representar objetos digitais de áudio, vídeo e outros formatos não previstos inicialmente, levando à realização de reuniões para revisar o padrão.

Em fevereiro de 2001, os membros da *Digital Library Federation* participaram de um *workshop* sobre o padrão de metadados usado pelo projeto *Making of America II* e decidiram pela criação de um novo formato de codificação (MCDONOUGH; MYRICK; STEDFELD, 2001). Nasceu, então, o padrão METS, que é baseado nos valiosos resultados teóricos e práticos do projeto *Making of America II* e que lhe asseguraram, desde o início, características de consistência, robustez e usabilidade.

3 METS

De acordo com seu registro na *National Information Standards Organization*, o *Metadata Encoding & Transmission Standard* (METS) é uma linguagem de marcação baseada em XML que provê uma estrutura capaz de registrar metadados descritivos, administrativos e estruturais relativos aos objetos de uma biblioteca digital (NISO, 2004). O padrão METS é expresso através de um XML Schema e um documento XML criado com base nesse padrão é denominado de *documento METS*.

O XML Schema de METS define um padrão para a codificação de documentos XML que podem conter uma relação de todos os arquivos ou, até mesmo, os próprios arquivos que compõem um ou diversos objetos digitais, juntamente com seus metadados descritivos, administrativos e estruturais. Com os metadados estruturais são registrados os relacionamentos existentes tanto entre os arquivos que compõem um objeto digital, como entre diferentes objetos digitais⁶. Esses relacionamentos podem ser do tipo hierárquico, ou em rede (comum em documentos com hipertexto).

6 Exemplos de relacionamentos entre os arquivos que compõem um objeto digital: arquivos de imagem e texto que formam uma página, páginas que pertencem a um capítulo, capítulos que formam um livro, dentre outros. Os relacionamentos entre diferentes objetos digitais podem ser ilustrados com livros que formam uma coleção, artigos e vídeos de palestras apresentados em um congresso, dentre outros.

Documentos METS simplificam tanto o gerenciamento do acervo de uma biblioteca digital, como o intercâmbio dos próprios objetos digitais e/ou de seus metadados entre instituições. Podem conter informações que orientem programas de computador (softwares) sobre como tratar ou apresentar alguns conteúdos do documento. Também proporcionam um formato comum para organizar a codificação dos metadados necessários ao processo de gerenciamento a longo prazo de objetos digitais. Isso permite que sejam empregados como qualquer um dos três tipos de pacotes de informação do padrão OAIS⁷ (vide seção “4 - METS e OAIS”).

O emprego de METS facilita a elaboração de um processo de interoperabilidade entre instituições e permite reduzir os custos de desenvolvimento, já que softwares desenvolvidos por uma instituição podem ser compartilhados entre todas as instituições adeptas desse padrão. Para Gartner (2002), METS preenche uma lacuna de padronização em relação aos metadados utilizados por bibliotecas digitais.

3.1 Responsáveis

No início de 2001, foi criado o padrão METS a partir de uma iniciativa da *Digital Library Federation*, detentora do *copyright* do *METS Schema* e do *METS Profile Schema* (NISO, 2004). A Biblioteca do Congresso Norte Americano é a agência mantenedora do padrão METS. Ela criou um sítio⁸ para hospedar as informações relativas a esse padrão.

Todas as alterações no padrão METS devem ser previamente analisadas e aprovadas por seu conselho editorial (NISO, 2004). A composição desse conselho pode ser encontrada no sítio desse padrão.

3.2 Usuários

A Universidade da Califórnia, em Berkeley, mantém um registro⁹ sobre as instituições que tornaram públicos os seus projetos de implantação do METS. O registro apresenta o estado em que cada projeto se encontra (em planejamento, em execução ou já concluído), uma breve descrição do projeto, sua documentação, uma relação das ferramentas utilizadas e o e-mail de pessoas que podem ser contatadas para troca de informações ou, até mesmo, de material como documentação e ferramentas criadas localmente. O registro tem a finalidade de criar uma comunidade entre as instituições que planejam implantar o METS e aquelas que já passaram por esse processo, para promover a troca de informações entre instituições e facilitar a difusão do padrão.

3.3 Estrutura

As informações desta seção sintetizam um abrangente estudo¹⁰ sobre o METS e têm como principal fundamento o XML Schema que define o METS e foi escrito por McDonough (2007). Também se utilizam algumas orientações encontradas na documentação de referência elaborada pelo METS Editorial Board (2007). O detalhamento apresentado limita-se ao necessário para evidenciar a riqueza de possibilidades de aplicações práticas desse padrão no contexto de uma biblioteca digital e promover sua difusão na comunidade.

7 Acrônimo do inglês *Open Archival Information System*.

8 Vide <http://www.loc.gov/standards/mets/>

9 Vide <http://www.loc.gov/standards/mets/mets-registry.html>

10 Estudo realizado pelo autor deste artigo durante a busca de um padrão de metadados expansível, capaz de conter outros tipos de metadados como partes integrantes dele mesmo, para uso em outra pesquisa.

A figura 1 apresenta a relação hierárquica entre os elementos XML que constituem o vocabulário desse padrão e, portanto, podem compor um documento METS. Recomenda-se ter a figura 1 à mão para melhor acompanhar e contextualizar as informações apresentadas nesta seção.

Na figura 1, observa-se que alguns elementos podem ser **obrigatórios**, **repetíveis** ou **subordinados a si mesmos**. Um elemento marcado como **obrigatório** deve constar de um documento METS sempre que o elemento ao qual é subordinado for codificado. Por exemplo, se o elemento *fileSec*¹¹, que é opcional, não constar de um documento METS, então o elemento *fileGrp* não deve ser codificado.

Muitos elementos são **repetíveis**, ou seja, podem aparecer múltiplas vezes. Por exemplo, se o elemento *fileSec*, que é opcional, constar de um documento METS, então deverá ter subordinado a ele um ou mais *fileGrp* que é um elemento obrigatório e repetível.

Os elementos *behaviorSec*, *fileGrp*, *file* e *div* podem ser **subordinados a si mesmos**. Por exemplo, o elemento *fileGrp* pode ter subordinado a ele, tanto o elemento *file*, como outro elemento *fileGrp*.

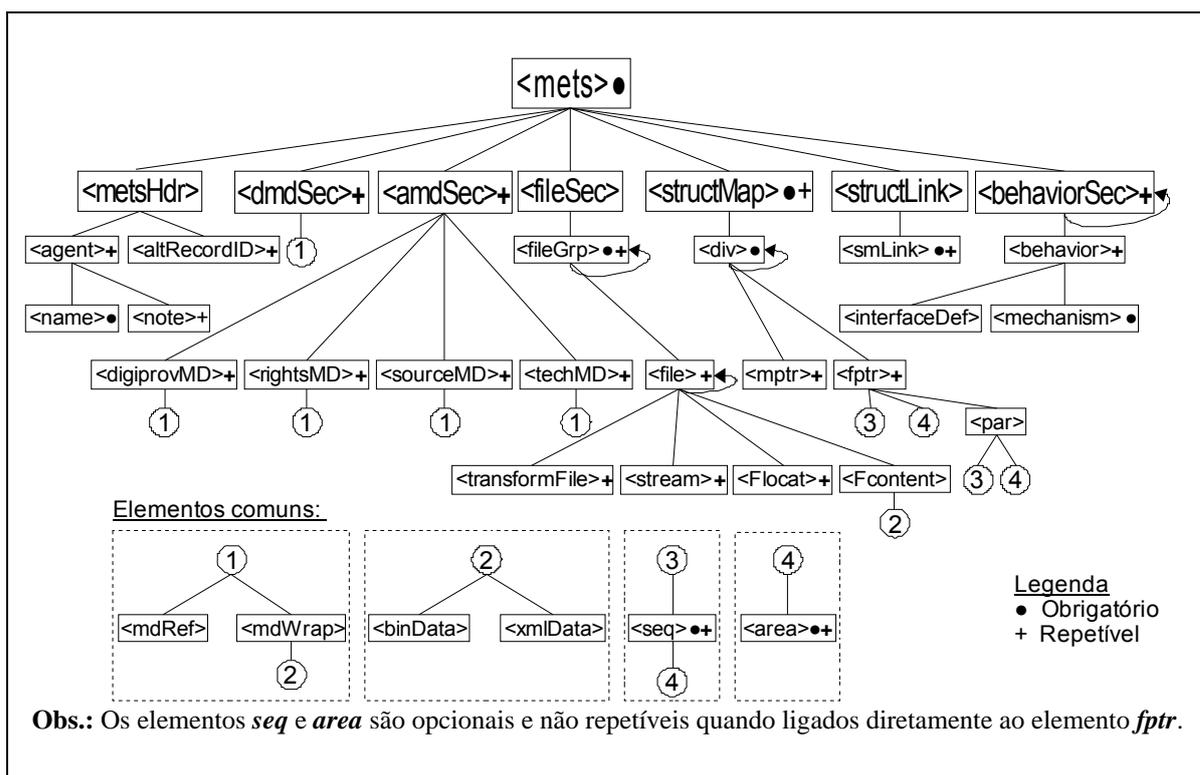


Figura 1 – Relação hierárquica entre os elementos XML que podem compor um documento METS. Fonte: elaboração própria a partir do XML Schema escrito por McDonough (2007), que define o METS.

De modo geral, cada elemento possui diversos atributos que são utilizados para qualificá-lo.

11 As menções feitas ao nome de cada elemento XML do padrão METS estão grafadas em itálico.

Esses atributos não estão representados na figura 1. Embora alguns dos exemplos apresentados adiante contenham atributos (alguns obrigatórios), este trabalho não relaciona nem descreve os atributos existentes para cada elemento XML de um documento METS. A relação completa dos atributos e uma breve descrição de seus domínios e significados podem ser encontradas no XML Schema escrito por McDonough (2007), que define o METS e é muito bem comentado.

Um documento METS pode ser formado por sete grandes seções, a saber: *metsHdr*, que identifica o próprio documento METS; *dmdSec*, que contém metadados descritivos; *amdSec*, que contém metadados administrativos; *fileSec*, que relaciona os arquivos que compõem o objeto digital descrito no documento METS; *structMap*, que representa a estrutura hierárquica entre os arquivos; *structLink*, que representa a relação entre nós da estrutura hierárquica e, por fim, *behaviorSec*, que associa o objeto digital a códigos executáveis (softwares) para interação com seu conteúdo. Essas grandes seções são detalhadas a seguir. Antes, faz-se necessário entender o papel dos elementos comuns.

3.3.1 Elementos comuns

Existem quatro grupos de elementos comuns na figura 1. O primeiro e o segundo grupos dão ao METS uma característica muito importante: a extensibilidade. METS não propõe um padrão próprio de metadados descritivos ou administrativos. O que ele oferece é uma estrutura capaz de registrar o padrão de metadados já utilizado por uma instituição, possibilitando o reaproveitamento do esforço empregado na montagem desses metadados.

No primeiro grupo de elementos comuns, *mdRef* é usado para indicar a localização de metadados externos ao documento METS, enquanto que *mdWrap* informa que os metadados fazem parte do próprio documento e serão codificados através dos elementos *binData* ou *xmlData*. No segundo grupo, *binData* é usado para armazenar um conteúdo (metadados ou arquivos) no padrão Base64, que foi descrito por Simon Josefsson (2003). *xmlData* é usado para armazenar conteúdos em formato XML.

O terceiro e quarto grupos de elementos comuns são usados para dar flexibilidade à forma como são feitas as referências aos arquivos que compõem um objeto digital. As referências podem ser apenas à parte de um arquivo (*area*), ou a múltiplas partes em seqüência (*seq*).

3.3.2 mets – Raiz do Documento

O elemento *mets* é a raiz de um documento METS. O exemplo 1 apresenta esse elemento com as declarações dos *namespaces* associados às estruturas XML utilizadas nos demais exemplos deste artigo.

<mets		
xmlns	=	'http://www.loc.gov/METS/'
xmlns:xlink	=	'http://www.w3.org/1999/xlink'
xmlns:dc	=	'http://purl.org/dc/elements/1.1/'
xmlns:xsi	=	'http://www.w3.org/2001/XMLSchema-instance'
xmlns:nisoimg	=	'http://www.loc.gov/mix/'
xmlns:oclc	=	'http://www.loc.gov/standards/premis/v1'
xmlns:rts	=	'http://cosimo.stanford.edu/sdr/metsrights/'
xsi:schemaLocation	=	
'http://www.loc.gov/METS/'		http://www.loc.gov/standards/mets/mets.xsd

http://www.w3.org/1999/xlink/	http://www.loc.gov/standards/mets/xlink.xsd
http://purl.org/dc/elements/1.1/	http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd
http://www.loc.gov/mix/	http://www.loc.gov/standards/mix/mix.xsd
http://www.loc.gov/standards/premis/v1	http://www.loc.gov/standards/premis/v1/PREMIS-v1-1.xsd
http://cosimo.stanford.edu/sdr/metsrights/	http://cosimo.stanford.edu/sdr/metsrights.xsd

Exemplo 1 – Elemento raiz de um documento METS que utiliza diversos esquemas de padrões de metadados.

3.3.3 *metsHdr* – METS Header (ou Cabeçalho METS)

Esta seção contém dados que descrevem o próprio documento METS. O elemento *agent* e seus subelementos (*name* e *note*) informam os responsáveis por cada papel (criador, editor, arquivista e outros) na elaboração desse documento.

Um objeto digital pode estar associado a um código de registro válido apenas para uma instituição, ou ao seu registro em outras entidades, tais como códigos ISBN¹² e ISSN¹³, dentre outros. O elemento *altRecordID*, representa essa informação.

```
<metsHdr CREATEDATE='2007-11-01T09:38:00'>
  <agent ROLE='CREATOR'>
    <name>Fulano Criador</name>
  </agent>
  <altRecordID TYPE='UNB'> CID0076543 </altRecordID>
</metsHdr>
```

Exemplo 2 – Cabeçalho de um documento METS informando quem o criou e o número de registro do objeto digital, interno à instituição.

3.3.4 *dmdSec* – Description Metadata Section (ou seção de metadados descritivos)

Contém metadados descritivos sobre o objeto digital como um todo, ou sobre suas partes integrantes. Conforme já mencionado, devido à característica de extensibilidade oferecida pelos elementos comuns *mdRef*, *mdWrap*, *binData* e *xmlData*, os metadados descritivos podem ser expressos por padrões como MARC, MODS, Dublin Core, TEI Header, EAD, VRA, FGDC, DDI, dentre outros, tanto no próprio documento METS, como em algum local externo ao documento.

No exemplo 3, vê-se como metadados Dublin Core podem ser utilizados para descrever o objeto digital. Optou-se por colocar os metadados dentro do próprio documento METS.

```
<dmdSec ID="DMD01">
  <mdWrap MDTYPE='DC' LABEL='Dublin Core'>
    <xmlData>
      <dc:title>Compreendendo o METS</dc:title>
      <dc:creator>Fulano Autor</dc:creator>
      <dc:subject>METS</dc:subject>
      <dc:subject>Biblioteca Digital</dc:subject>
      <dc:subject>OAIS</dc:subject>
      <dc:publisher>CID-UnB</dc:publisher>
      <dc:language>pt-BR</dc:language>
      <dc:date>2007-10-27</dc:date>
```

12 Acrônimo do Inglês International Standard Book Number. Vide <http://www.isbn-international.org/>

13 Acrônimo do Inglês International Standard Serial Number. Vide <http://www.issn.org/>

```
</xmlData>
</mdWrap>
</dmdSec>
```

Exemplo 3 – Metadados descritivos em um documento METS usando o padrão Dublin Core.

3.3.5 *amdSec* – Administrative Metadata Section (ou seção de metadados administrativos)

Essa seção funciona como um agrupamento de quatro tipos de metadados de caráter administrativo:

– *techMD* (*technical metadata*) – Metadados técnicos relativos ao objeto digital, tais como formato do arquivo, resolução, tamanho e outros. Um exemplo desse tipo de metadado é o *NISO Metadata for Images in XML* (NISO MIX)¹⁴.

– *rightsMD* (*rights metadata*) – Metadados sobre direitos de propriedade intelectual, tais como restrições de uso, observações e outros. Um exemplo desse tipo de metadado é METSRights que Barlas (2006) define como um XML Schema que estende um documento METS com metadados sobre direitos de propriedade intelectual relativos a um objeto digital.

– *sourceMD* (*source metadata*) – Metadados relativos ao original, não digital, a partir do qual foi criado o objeto digital.

– *digiprovMD* (*digital provenance metadata*) – Metadados sobre a procedência do objeto digital. Caso o objeto digital tenha sido obtido a partir de outro objeto digital, essa seção registrará um histórico das transformações que foram aplicadas desde o primeiro objeto, até que se obtivesse o atual. Como exemplo, considere-se um original que foi digitalizado gerando um arquivo de imagem no formato TIFF e este, por sua vez, foi convertido em uma imagem no formato JPEG. Então, os metadados de procedência desta imagem JPEG devem informar que ela não foi obtida diretamente do original e sim da imagem TIFF. Isso permite que o usuário desse documento determine se esse objeto digital ainda tem condições de espelhar características específicas dos originais. Um exemplo desse tipo de metadado é apresentado pelo PREMIS Working Group (2005) e foi desenvolvido em conjunto pela OCLC e RLG¹⁵.

```
<amdSec ID="AMD01">
  <techMD ID="TMD01">
    <mdWrap MDTYPE='NISOIMG' LABEL='NISO'>
      <xmlData> ...metadados de informações técnicas </xmlData>
    </mdWrap>
  </techMD>
  <rightsMD ID="RMD01">
    <mdWrap MDTYPE='OTHER' LABEL='METSRights'>
      <xmlData> ...metadados sobre propriedade intelectual </xmlData>
    </mdWrap>
  </rightsMD>
  <sourceMD ID="SMD01">
    <mdWrap MDTYPE='DC' LABEL='Dublin Core'>
      <xmlData> ...metadados de origem </xmlData>
    </mdWrap>
```

14 Vide <http://www.loc.gov/standards/mix/>

15 Vide <http://www.oclc.org/research/projects/pmwg/>

```

</sourceMD>
<digiprovMD ID="DPMD01">
  <mdWrap MDTYPE='OTHER' LABEL='oclcprov'>
    <xmlData> ...metadados de procedência </xmlData>
  </mdWrap>
</digiprovMD>
</amdSec>

```

Exemplo 4 – Declaração de diversos metadados administrativos em um documento METS.

As declarações desses metadados podem estar externas ou internas ao documento. O exemplo 4 ilustra a estrutura XML necessária à seção de metadados administrativos para abrigar quatro tipos de metadados diferentes dentro do próprio documento METS.

3.3.6 *fileSec* – File Section (ou Seção de Arquivo)

Um objeto digital pode ser formado por múltiplos arquivos de imagem, texto, áudio, vídeo, entre outros. O propósito desta seção é relacionar todos os arquivos que compõem o objeto digital e onde podem ser encontrados.

Os arquivos devem ser agrupados em *fileGrp* que podem ser subordinados a si mesmos, de forma análoga às pastas/diretórios de um sistema de arquivos. As declarações de arquivo, *file*, também podem ser subordinadas a si mesmas. Isso é útil para representar casos como o de grandes arquivos que são fracionados em múltiplos arquivos menores. Tanto *fileGrp* como *file* possuem atributos que permitem associá-los a metadados administrativos, *amdSec*.

O local onde se encontra um arquivo do objeto digital é indicada pelo elemento *FLocat*. Um arquivo pode ser embutido em um documento METS utilizando o elemento *FContent*.

Pode-se associar um tipo de transformação, *transformFile*, a ser aplicada ao arquivo, como no caso de arquivos criptografados que devem ser decifrados antes de serem exibidos. É possível, também, indicar que um arquivo de filme possui um canal de áudio e outro de vídeo, cada qual com seus próprios metadados, através do elemento *stream*.

No exemplo 5, no primeiro *fileGrp* são declarados três arquivos de imagem obtidos por digitalização das páginas de um documento original e, no segundo *fileGrp* três arquivos contendo o OCR de cada uma das imagens.

```

<fileSec>
  <fileGrp USE='Imagem'>
    <file ID='IMG0001' MIMETYPE='image/tiff'>
      <FLocat xlink:href='http://cid.unb.br/0001.jpg' LOCTYPE='URL'/>
    </file>
    <file ID='IMG0002' MIMETYPE='image/tiff'>
      <FLocat xlink:href='http://cid.unb.br/0002.jpg' LOCTYPE='URL'/>
    </file>
    <file ID='IMG0003' MIMETYPE='image/tiff'>
      <FLocat xlink:href='http://cid.unb.br/0003.jpg' LOCTYPE='URL'/>
    </file>
  </fileGrp>
  <fileGrp USE='Ocr de Imagem'>
    <file ID='OCR0001' MIMETYPE='text/plain'>
      <FLocat xlink:href='http://cid.unb.br/0001.txt' LOCTYPE='URL'/>
    </file>
  </fileGrp>
</fileSec>

```

```

        </file>
        <file ID='OCR0002' MIMETYPE='text/plain'>
            <FLocat xlink:href='http://cid.unb.br/0002.txt' LOCTYPE='URL'/>
        </file>
        <file ID='OCR0003' MIMETYPE='text/plain'>
            <FLocat xlink:href='http://cid.unb.br/0003.txt' LOCTYPE='URL'/>
        </file>
    </fileGrp>
</fileSec>

```

Exemplo 5 – Declaração de três arquivos de imagem e seus arquivos de OCR.

3.3.7 *structMap* – Structural Map (ou Mapa Estrutural)

A seção de mapa estrutural fornece os meios para representar uma estrutura hierárquica que organiza outros documentos METS, ou os arquivos (*file*) informados em *fileSec*. Essa estrutura hierárquica é denotada principalmente pelo elemento *div*, que pode ser subordinado a si mesmo e tem uma finalidade de organização semelhante às pastas (diretórios) em uma estrutura de arquivos.

```

<structMap TYPE="fisico">
    <div ADMID="AMD01" DMDID="DMD01">
        <div TYPE="pagina" ORDER="1" xlink:label='PAG0001'>
            <fptr FILEID="IMG0001"/>
            <fptr FILEID="OCR0001"/>
        </div>
        <div TYPE="pagina" ORDER="2">
            <fptr FILEID="IMG0002"/>
            <fptr FILEID="OCR0002"/>
        </div>
        <div TYPE="pagina" ORDER="3" xlink:label='PAG0003'>
            <fptr FILEID="IMG0003"/>
            <fptr FILEID="OCR0003"/>
        </div>
    </div>
</structMap>

```

Exemplo 6 – Declaração de um mapa estrutural.

O elemento *mptr* permite fazer referência a outros documentos METS, enquanto que *fptr* é usado para fazer referência a arquivos (elementos *file*). Essa referência pode ser bem flexível devido ao uso dos elementos comuns *area* e *seq*, descritos anteriormente. Também pode ser informado que dois arquivos possuem um paralelismo, *par*, como no caso de um arquivo de vídeo e um arquivo de legenda que devem ser tratados (exibidos) simultaneamente.

Um mesmo documento METS pode conter mais de um mapa estrutural. Isso permite declarar diferentes estruturas hierárquicas para um mesmo conjunto de arquivos, tais como a organização lógica ou física que eles possuem. A figura 1 mostra que das sete grandes seções, apenas a de mapa estrutural (*structMap*) é obrigatória.

O mapa estrutural apresentado no exemplo 6 é bem simples e representa a estrutura física de um objeto digital com três páginas, sendo cada uma composta por uma imagem e um texto obtido por OCR.

3.3.8 *structLink* – Structural Map Linking (ou Vinculação de Mapa Estrutural)

A seção de vinculação de mapa estrutural especifica “hyperlinks” entre diferentes divisões de um mapa estrutural (*structMap*). Enquanto *structMap* apresenta somente relações **hierárquicas** entre as divisões estruturais (*div*), *structLink* permite declarar a referência de uma divisão estrutural (*div*) à outra, estejam elas ligadas hierarquicamente, ou não. Isso permite a representação das relações **não hierárquicas** existentes no objeto digital. Esse tipo de relação é comum em documentos com hipertexto, onde uma página de um capítulo pode conter “hyperlinks” para outras páginas em outros capítulos ou, até mesmo, em outros documentos.

O elemento *smLink* indica quais os dois nós (elementos *div*) do mapa estrutural que estão ligados e em que sentido se dá a ligação. O exemplo 7 mostra como estabelecer um “hyperlink” da primeira com a última página do exemplo anterior:

```
<structLink>
  <smLink xlink:from='PAG0001' xlink:to='PAG0003'/>
</structLink>
```

Exemplo 7 – “Hyperlink” entre duas divisões estruturais.

3.3.9 *behaviorSec* – Behavior Section (ou Seção de Comportamento)

A seção de comportamento permite associar código executável (um programa ou um serviço) a um conteúdo do documento METS. Isso traz ao conteúdo uma característica dinâmica, um comportamento.

O elemento *behavior* indica a qual conteúdo do documento se deseja associar um código executável. *mechanism* indica onde este código pode ser encontrado: se no próprio documento ou em algum lugar externo a ele como, por exemplo, um “web service”¹⁶.

A descrição das funcionalidades que podem ser executadas pelo código, ou seja, a definição de sua interface externa (métodos e propriedades) pode ser encontrada no local referenciado pelo elemento *interfaceDef*, que pode estar no próprio documento METS ou externo a ele.

Como exemplo, considere-se o caso em que um conteúdo deva ser apresentado para o usuário, mas seu formato de armazenamento não é diretamente adequado para exibição. Para resolver esse problema, pode ser criado um *behavior* responsável pela exibição desse conteúdo, ou seja, um procedimento para transformar esse conteúdo em um formato próprio para exibição.

Isso encerra a apresentação das características mais relevantes sobre cada uma das seções de um documento METS. A seguir, discute-se um recurso criado para descrever de forma bem detalhada as orientações necessárias para criação e processamento de documentos METS, facilitando o intercâmbio destes documentos entre diferentes instituições.

3.4 Profiles

Os *profiles* são usados para simplificar tanto a implantação do METS como o compartilhamento de objetos digitais e seus metadados. Para que uma instituição utilize o METS com um determinado tipo de acervo, dentre outras ações, deve definir qual a melhor

¹⁶ Trata-se de uma forma de prover serviços via web, não para usuários, mas entre programas de computador. Vide <http://www.w3.org/2002/ws/>, para mais detalhes.

estrutura para representar os elementos do acervo, os metadados mais adequados para descrevê-los, os elementos opcionais de METS que devem ser utilizados e os domínios de valores para seus atributos. Esse detalhamento de implantação pode ser relatado em um formato pré-definido de documento XML: o *Profile*.

O Conselho Editorial do METS recomenda que cada *profile* seja submetido a um processo formal para seu registro. Nesse processo, o Conselho faz uma avaliação do *profile* que, se aprovado, passa a integrar uma lista¹⁷ de *profiles* recomendados. Tal processo tem a finalidade de estabelecer práticas comuns entre usuários do METS, já que os *profiles* registrados são de acesso público. Os novos adeptos desse padrão são estimulados a utilizar *profiles* já aprovados, ao invés de criarem os seus próprios.

A aplicação de METS não está restrita a bibliotecas digitais. A relação de METS com o modelo Open Archival Information System é um significativo exemplo disso.

4 METS e OAIS

Devido à semelhança entre as siglas, é comum a confusão entre OAI (*Open Archives Initiative*) e OAIS (*Open Archival Information System*). A primeira é definida por Lagoze e Sompel (2001) como uma iniciativa que busca desenvolver e promover soluções de interoperabilidade que objetivam facilitar uma disseminação eficiente de conteúdo. A segunda foi elaborada pelo Comitê Consultivo para Sistemas de Dados Espaciais da NASA (CCSDS), tornando-se o padrão ISO 14721 em 2002.

O OAIS consiste em uma organização de pessoas e de sistemas, que têm a responsabilidade de preservar a informação e torná-la disponível para uma determinada comunidade (CCSDS, 2002). A figura 2 apresenta o fluxo de informação entre seus produtores, uma estrutura OAIS e seus consumidores, conforme definições do CCSDS (2002).

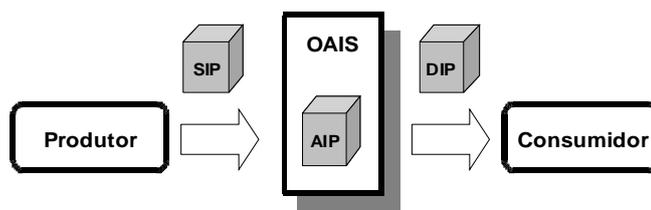


Figura 2 – Macro visão da interação entre entidades de um ambiente OAIS.

Fonte: elaboração própria a partir das definições do CCSDS (2002).

Produtores são pessoas ou sistemas responsáveis por fornecer a informação a ser preservada. Essa informação deve ser submetida na forma de pacotes denominados SIP¹⁸. O ambiente OAIS valida os pacotes SIP e, a partir deles, gera pacotes denominados AIP¹⁹ com um conjunto de informações adequados à preservação. Consumidores são pessoas ou sistemas que utilizam os serviços do ambiente OAIS para localizar e recuperar uma informação por ele preservada. As requisições feitas pelos consumidores são respondidas com pacotes

17 Vide <http://www.loc.gov/standards/mets/mets-registered-profiles.html>

18 Acrônimo do inglês *Submission Information Package*.

19 Acrônimo do inglês *Archival Information Package*.

denominados DIP²⁰.

O modelo de referência elaborado pelo CCSDS (2002) define quatro níveis de interação entre ambientes OAIS, a saber: Independentes, Cooperados, Federados e com Recursos Compartilhados. Quanto maior o nível de interação, maior a necessidade de padronização dos pacotes empregados. Apesar disso, o modelo OAIS não detalha o conteúdo de cada pacote e deixa que as partes envolvidas entrem em um entendimento quanto a isso. Portanto, podem existir diferenças significativas entre os pacotes adotados por dois ambientes OAIS.

Um documento METS pode desempenhar o papel de um pacote SIP, AIP ou DIP dependendo do conjunto de informações nele contido (BEEDHAM, 2005). O emprego de METS com essa finalidade pode simplificar a tarefa de padronização desses pacotes entre múltiplos ambientes OAIS. Isso representa um importante passo para viabilizar a montagem de ambientes OAIS com níveis de cooperação mais elevados.

O CCSDS (2002) define quatro tipos de informações como necessárias para preservação em um ambiente OAIS: Procedência, Contexto, Referência e Fixidez. Quando o METS é usado como uma forma de empacotar informações para fazer transferências, ele pode representar um pacote SIP ou DIP. Quando contiver um conjunto de informações suficientes para preservação, pode ser usado como um pacote AIP.

5 OUTROS PADRÕES

METS surgiu da necessidade que a comunidade ligada às bibliotecas teve de criar metadados adequados para tratar objetos digitais. Entretanto, outras comunidades buscaram suas próprias soluções. Algumas dessas iniciativas estão relacionadas a seguir:

– **IMS-CP** – Criado pela comunidade relacionada à área de educação. Define um conjunto de estruturas padronizadas usadas para o intercâmbio de material referente à educação (IMS, 2004).

– **MPEG-21 Digital Item Declaration Language (DIDL)** – Foi criado pelo *Moving Picture Experts Group* (MPEG), que desenvolve padrões para representar áudio e vídeo digitais. Segundo Bekaert e Sompel (2005), o MPEG-21 define um conjunto de ferramentas que permite distribuir e consumir arquivos multimídia entre usuários (pessoas ou sistemas que interagem com o MPEG-21).

– **XFDU - XML Formatted Data Unit** – Foi criado pelo Comitê Consultivo para Sistemas de Dados Espaciais da NASA (CCSDS) e permite criar um único pacote contendo dados (inclusive software) e metadados. De acordo com CCSDS (2004), o objetivo do XFDU é facilitar a transferência e o arquivamento de todo tipo de informação digital no âmbito da agência espacial.

– **SCORM (Sharable Content Object Reference Model)** – O Departamento de Defesa dos Estados Unidos e o Escritório de Políticas de Ciência e Tecnologia da Casa Branca deram início em 1997 a uma iniciativa de aprendizagem distribuída. Essa iniciativa desenvolveu o

20 Acrônimo do inglês *Dissemination Information Package*.

SCORM para estimular a criação de conteúdos de aprendizagem reutilizáveis (uma aplicação do conceito de orientação a objetos direcionada à educação). O SCORM apresenta um modelo para agregação de conteúdo que permite a criação de pacotes com conteúdo educacional, metadados e um detalhamento de como percorrer (navegar) pelo conteúdo (ADL, 2006).

– **RAMLET** – É uma iniciativa do Institute of Electrical and Electronics Engineers, IEEE, e apresenta um modelo para agregação de recursos de aprendizagem, educacionais e de treinamento (IEEE, 2005).

6 CONSIDERAÇÕES FINAIS

O METS foi desenvolvido a partir das avaliações técnicas produzidas pelo projeto *Making of America II* sobre as melhores práticas que deveriam ser adotadas para a montagem de bibliotecas digitais. Como acontece com qualquer padrão, sua importância está diretamente associada ao seu uso. Instituições renomadas como as bibliotecas digitais das universidades de Harvard, da Califórnia e de Oxford, além da Online Computer Library Center (OCLC) e da biblioteca do Congresso Norte Americano, dentre outras, já adotaram esse padrão para algumas de suas coleções. Muitas outras instituições ainda o estão avaliando.

Os problemas associados à pesquisa e recuperação de objetos digitais, ao intercâmbio entre diferentes instituições e à diversidade de esquemas e tipos de metadados usados para a preservação digital podem ser bastante minimizados com a maciça adoção desse padrão.

Por fim, a importância desse padrão foi claramente expressa por Gartner (2003) ao dizer que o METS vem oferecer a perspectiva de estabelecer um padrão coerente e flexível que trará às bibliotecas digitais algumas das vantagens que o registro MARC trouxe à sua contraparte tradicional (bibliotecas não digitais).

AGRADECIMENTO

Agradecemos a João Alberto de Oliveira Lima, doutor em Ciência da Informação, por suas valiosas orientações durante todas as fases da elaboração desse artigo.

REFERÊNCIAS

ADL. **Content Aggregation Model (CAM)**. Virginia, 2006. Disponível em: <<http://www.adlnet.gov/downloads/DownloadPage.aspx?ID=237>>. Acesso em: 07 mai. 2008.

BARLAS, Chris. **Digital Rights Expression Languages (DREs)**. [S.l.]:JISC, 2006. Disponível em: <<http://www.jisc.ac.uk/media/documents/techwatch/tsw0603.pdf>>. Acesso em: 07 mai. 2008.

BEEDHAM, Hilary et al. **Assessment of UK Data Archive and The National Archives compliance with OAIS/METS**. [S.l.]: JISC, 2005. Disponível em: <http://www.jisc.ac.uk/whatwedo/programmes/programme_preservation/programme_404/project_oais.aspx>. Acesso em: 07 mai. 2008.

BEKAERT, Jeroen; SOMPEL, Herbert Van de. **Representing Digital Assets using MPEG-21 Digital Item Declaration**. [S.l.], 2005. Disponível em: <<http://arxiv.org/abs/cs/0508065>>.

Acesso em: 07 mai. 2008.

CCSDS. Reference Model for an Open Archival Information System (OAIS).

Washington:NASA, 2002. Disponível em:

<<http://public.ccsds.org/publications/archive/650x0b1.pdf>>. Acesso em: 07 mai. 2008.

_____. XML Formatted Data Unit (XFDU) Structure and Construction Rules.

Washington:NASA, 2004. Disponível em:

<<http://sindbad.gsfc.nasa.gov/xfdu/pdfdocs/iprwbv2a.pdf>>. Acesso em: 07 mai. 2008.

CUNHA, Murilo Bastos da. Biblioteca digital: bibliografia internacional anotada. **Ciência da Informação**, V. 26, N. 2 p. 195-213. Ibict, 1997. Disponível em:

<<http://www.ibict.br/cionline/viewarticle.php?id=442&layout=abstract>>. Acesso em: 07 mai. 2008.

DLF. Structural, technical, and administrative metadata standards. A discussion document. [S.l.]: CLIR, 2000. Disponível em:

<<http://www.diglib.org/standards/stamdfame.htm>>. Acesso em: 10 dez. 2006.

GARTNER, Richard. **METS: Metadata Encoding and Transmission Standard.** [S.l.]:

JISC, 2002. Disponível em: <http://www.jisc.ac.uk/uploaded_documents/tsw_02-05.pdf>.

Acesso em: 07 mai. 2008.

_____. METS: Implementing a metadata standard in the digital library. [S.l.], 2003.

Disponível

em:<http://www.iatul.org/doclibrary/public/Conf_Proceedings/2003/GARTNER_fulltext.pdf

>. Acesso em: 07 mai. 2008.

HURLEY, Bernard J. et al. **The Making of America II Testbed Project: A Digital Library Service Model.** Washington: CLIR, 1999. Disponível em:

<<http://www.clir.org/pubs/reports/pub87/pub87.pdf>>. Acesso em: 07 mai. 2008.

IEEE. Developing a reference model for resource aggregation for learning, education, and training. Nova Iorque, 2005. Disponível em:

<http://ieeeltsc.org/wg11CMI/ramlet/Pub/RAMLET_project_description.pdf>. Acesso em: 07 mai. 2008.

IMS. IMS Content Packaging Information Model. [S.l.]:IMS, 2004. Disponível em:

<http://www.msglobal.org/content/packaging/cpv1p1p4/imscp_infov1p1p4.html>. Acesso em: 07 mai. 2008.

LAGOZE, Carl; SOMPEL, Herbert Van de. The Open Archives Initiative: Building a low-barrier interoperability framework. [S.l.]:JCDL, 2001. Disponível em:

<<http://www.openarchives.org/documents/jcdl2001-oai.pdf>>. Acesso em: 07 mai. 2008.

MCDONOUGH, Jerome. METS Schema 1.7.[S.l.]:LOC, 2007. Disponível em:

<<http://www.loc.gov/standards/mets/mets.xsd>>. Acesso em: 07 mai. 2008.

MCDONOUGH, Jerome; MYRICK, Leslie; STEDFELD, Eric. Report on The Making of America II DTD - Digital Library Federation Workshop. [S.l.]: CLIR, 2001. Disponível em:

<<http://www.diglib.org/standards/metssum.pdf>>. Acesso em: 07 mai. 2008.

METS Editorial Board. METS: Primer and Reference Manual. [S.l.]:LOC, 2007.

Disponível em:

<<http://www.loc.gov/standards/mets/METS%20Documentation%20final%20070930%20msw.pdf>>. Acesso em: 30 jul. 2008.

NISO. **Metadata Encoding & Transmission Standard (METS)**. [S.l.], 2004. Disponível em: <<http://www.niso.org/publications/registrations/approved/METSregweb.pdf>>. Acesso em: 31 jul. 2008.

PREMIS WORKING GROUP. **Data Dictionary for Preservation Metadata**. [S.l.]: OCLC/RLG, 2005. Disponível em: <<http://www.oclc.org/research/projects/pmwg/premis-final.pdf>>. Acesso em: 07 mai. 2008.

SIMON JOSEFSSON. **RFC 3548 - The Base16, Base32, and Base64 Data Encodings**. [S.l.]: IETF, 2003. Disponível em: <<http://www.ietf.org/rfc/rfc3548>>. Acesso em: 07 mai. 2008.

UC BERKELEY LIBRARY. **The Making of America II - Proposal to DLF**. Berkeley, 1997a. Disponível em: <<http://sunsite.berkeley.edu/moa2/planning.pdf>>. Acesso em: 07 mai. 2008.

_____. **The Making of America II Testbed Project – Proposal to NEH**. Berkeley, 1997b. Disponível em: <<http://sunsite.berkeley.edu/moa2/moaproposal.html>>. Acesso em: 07 mai. 2008.

ABSTRACT

The Metadata Encoding & Transmission Standard (METS) is an XML Schema created under the sponsorship of the Digital Library Federation and maintained by the Library of Congress which allows to create a rich structure capable to hold not only the multiple types of metadata used to describe the assets of a digital library, but also the own digital objects, be for management purposes, be to allow an interchange between different institutions. This article presents the motivation problems for its creation and its historical origin in the Making of America II project. Describes the purpose of each XML element that forms the vocabulary of this standard and shows the usage of some of them with XML coded examples, which requires a basic knowledge of that language for better understanding. Expresses its relation with the Open Archival Information System model as a meaningful example of its use and, finally, mentions other standards created with similar purposes.

KEYWORDS: Metadata. Digital library. Digital preservation. METS. Metadata Encoding & Transmission Standard.

Originals recebidos em: 27/05/2008

Texto aprovado em: 02/09/2008