



**SENADO FEDERAL
UNIVERSIDADE DO LEGISLATIVO BRASILEIRO
UNILEGIS**

EDGAR BENÍCIO ROSA

Preservação digital de áudios da Rádio Senado

Brasília – DF
2009



EDGAR BENÍCIO ROSA

Preservação digital de áudios da Rádio Senado

Trabalho final apresentado como requisito para aprovação no curso de pós-graduação *lato sensu* em Administração Legislativa realizado pela Universidade do Legislativo Brasileiro e Universidade Federal do Mato Grosso do Sul como requisito para obtenção do título de especialista em Administração Legislativa.

Orientador: João Alberto de Oliveira Lima

Brasília
2009

Preservação digital de áudios da Rádio Senado

Trabalho de Conclusão do Curso de Especialização em Administração Legislativa promovido pela Universidade do Legislativo Brasileiro e Universidade Federal do Mato Grosso do Sul no 1º semestre de 2009.

Aluno: Edgar Benício Rosa

Banca Examinadora:

João Alberto de Oliveira Lima

Ernesto Wilhelms Neto

Brasília, 04 de agosto de 2009

AGRADECIMENTOS

Ao amigo e orientador João Alberto de Oliveira Lima pela sua valiosa colaboração.

A todos que direta ou indiretamente, colaboraram para a execução deste trabalho.

"Para conquistar coisas importantes,
devemos não apenas agir mas também sonhar,
não apenas planejar mas também acreditar ."
(Anatole France)

RESUMO

A preservação de objetos digitais é um problema que as instituições tem se deparado a partir do advento da migração do seu acervo físico para o meio digital e da geração de novas informações no formato digital. O volume crescente das informações digitais que tem necessidade de ser preservadas bem a grande quantidade de informações igualmente armazenadas sem necessidade (lixo eletrônico) tem provocado a pesquisa e a busca por uma solução que garantisse a preservação, recuperação e disponibilização dos acervos digitais com segurança. A Rádio Senado enfrenta este problema, na medida em que, diariamente, são gerados vários produtos digitais em texto e áudio que são armazenados sem a adequada indexação e sem uma política definida de conservação deste acervo. O estudo *Preservação Digital de Áudios da Rádio Senado* pretende analisar os requisitos necessários para a preservação de objetos digitais a longo prazo, em especial áudio e texto. Analisaremos o modelo *Open Archive Information System* (OAIS) que tem se destacado como a proposta mais promissora a ser seguida, sendo nele previsto e tratado todas as etapas envolvidas em sistema de preservação digital. Criamos um exemplo, que se aplica à realidade da Rádio Senado, para a etapa inicial do processo de preservação utilizando o esquema *Metadata Encoding & Transmission Standard* (METS) para os objetos digitais complexos.

Palavras-chave: Áudio digital; Preservação digital; Metadados; OAIS; METS.

LISTA DE TABELAS

Tabela 1 – Diferentes tipos de metadados e suas funções.....	21
Tabela 2 – Atributos e características dos metadados	22
Tabela 3 – Agentes envolvidos no processo de preservação.....	41
Tabela 4 – Elementos básicos no Dublin Core.....	43
Tabela 5 – Elementos do esquema de metadados para áudio.....	44
Tabela 6 – Esquema de metadados de direitos autorais	45

LISTA DE FIGURAS

Figura 1 – Mais bits resulta em maior faixa dinâmica	15
Figura 2 – Digitalização de um sinal de áudio	15
Figura 3 – Layout interno de um arquivo mp3 com ID3v1 e ID3v2.....	23
Figura 4 – Tela do software WINAMP mostrando metadados internos ao arquivo de áudio.	24
Figura 5 – Tela de edição dos metadados ID3V2.....	24
Figura 6 – O ambiente OAIS	32
Figura 7 – Obtendo informações a partir dos dados armazenados	32
Figura 8 – Relacionamentos e conceitos do pacote de informação	33
Figura 9 – Entidades funcionais do OAIS e os pacotes de informação.....	34
Figura 10 – Diagrama funcional do DSPACE	36
Figura 11 – Estrutura interna do Fedora.....	37
Figura 12 – Estrutura de um documento METS.....	39
Figura 13 – Raiz de um documento METS	40
Figura 14 – Metadados de cabeçalho	40
Figura 15 – Metadados descritivos Dublin Core.....	42
Figura 16 – Exemplo de metadados técnicos de um arquivo de áudio no formato MP3	43
Figura 17 – Exemplo de metadados de direitos autorais	45
Figura 18 – Pacote contendo áudio e texto.....	46
Figura 19 – Pacote contendo somente áudio	46
Figura 20 – Exemplo de Seção de arquivos	47
Figura 21 – Exemplo de Mapa Estrutural.....	47
Figura 22 – Exemplo de Mapa Estrutural com referência a outros objetos METS.....	48
Figura 23 – Relacionamentos existentes no exemplo criado no padrão METS	49
Figura 24 – Ligações existentes em um documento METS	54

ABREVIATURAS

AIP – Archival Information Package

CCSDS – Consultative Committee for Space Data Systems

CD – Compact Disc

DC – Dublin Core

DIP – Dissemination Information Package

FM – Modulação em Frequência

HTML – Hypertext Markup Language

METS – Metadata Encoding & Transmission Standard

MP3 – MPEG-1/2 Audio Layer 3

MPEG – Moving Picture Experts Group

NASA – National Aeronautics and Space Administration

OAIS – Open Archive Information System

OC – Ondas Curtas

PCM – Pulse Code Modulation

PDF – Portable Document Format

SICON - Sistema de Informação do Congresso Nacional.

SIP – Submission Information Package

WAV – Waveform audio format

XML – Extensible Markup Language

SUMÁRIO

LISTA DE TABELAS

LISTA DE FIGURAS

ABREVIATURAS

1 INTRODUÇÃO.....	11
1.1 Objetivo Geral	12
1.2 Objetivos Específicos	12
1.2 Organização do trabalho	13
2 PADRÕES DE FORMATOS PARA OBJETOS DIGITAIS.....	14
2.1 FORMATOS PARA ÁUDIOS DIGITAIS	14
2.2 FORMATOS PARA TEXTOS DIGITAIS	16
2.2.1 Arquivos TXT	16
2.2.2 Arquivos DOC.....	17
2.2.3 Arquivos PDF.....	17
2.3 FORMATOS PARA COMPRESSÃO SEM PERDAS	18
3 METADADOS	19
3.1 Tipos de metadados	19
3.2 Metadados existentes em arquivos de áudio.....	23
4 PRESERVAÇÃO DIGITAL	25
4.1 Estratégias da preservação digital.....	27
4.1.1 Estratégias estruturais	27
4.1.2 Estratégias operacionais	27
4.2 Segurança da Informação	28
5 MODELO OAIS (Open Archive Information System).....	31
5.1 O ambiente OAIS	31
5.2 Os Pacotes de Informação	32
5.3 A entidade Produtor.....	35
5.4 Exemplos de sistemas utilizando o modelo OAIS.....	35
6 IMPLEMENTAÇÃO DE UM EXEMPLO.....	38
6.1 METS.....	38
7 CONCLUSÃO.....	50
7.1 Estudos Futuros	50
REFERÊNCIAS BIBLIOGRÁFICAS	52
ANEXO A	54
ANEXO B	55
ANEXO C	61

1 INTRODUÇÃO

A Rádio Senado foi criada com o objetivo de transmitir o áudio das reuniões das comissões e das sessões plenárias do Senado Federal e do Congresso Nacional. Também divulga as demais atividades do Senado, incluindo as ações de sua Presidência e dos senadores, levando à população todas as discussões e projetos votados na Casa.

Além da transmissão ao vivo das principais atividades do Senado Federal, as informações resumidas sobre as decisões, debates, discussões e propostas dos senadores são apresentadas em forma de reportagem pela equipe de jornalismo. Também são produzidas reportagens e entrevistas sobre fatos que não envolvem diretamente a atividade legislativa, mas que merecem análise e debate pelos senadores e interessam à sociedade. Conta ainda com produção própria e de terceiros de programas culturais.

A Rádio Senado tem em sua estrutura os setores de produção, arquivo, técnica, operações, redação, locução, administração, além dos núcleos de Ondas Curtas e de RadioAgência. Dispõe ainda de quatro estúdios, sendo dois deles utilizados para as transmissões ao vivo (FM e OC) e os demais para as gravações de programas e matérias jornalísticas. A sua programação gravada em formato digital é veiculada de forma inteiramente automatizada.

As matérias jornalísticas são produzidas utilizando-se o software Microsoft Word para a digitação dos textos e, em seguida, elas são lidas “ao vivo” nas emissoras ou gravadas em computador para irem “ao ar” posteriormente. As gravações de áudio são realizadas no formato digital (mp3) utilizando-se o software Adobe Audition. Todo o material gerado é armazenado em um servidor com acesso “para leitura” a todos os funcionários da emissora, com acesso para “escrita” para a área técnica e acesso “total” para a área de suporte, sem indexação, sem uma ligação eficiente entre os arquivos de texto e áudio, o que dificulta a sua localização e a proteção dos dados armazenados. O material produzido pela RadioAgência fica também disponível para o público em geral, através da Internet, mas seu acervo é reduzido devido à falta de espaço para armazenamento e não tem um eficiente mecanismo de busca. Todo este material digital produzido necessita ser armazenado e disponibilizado seguindo normas eficientes de segurança de dados de forma a não sofrer qualquer tipo de alteração em seu conteúdo, preservando a história e a veracidade das informações armazenadas. O armazenamento digital observando os requisitos de segurança de informação

aliados à indexação dos dados e à utilização de metadados proporcionará um acesso seguro e eficiente a toda base de dados gerada pela Rádio Senado.

Como vantagens do armazenamento digital podemos destacar:

- economia de espaço físico;
- conservação do material de consulta (uma vez que não há desgaste devido ao manuseio deste material);
- sobretudo a facilidade de acesso, possibilitando que o consumidor da informação se encontre em qualquer local com acesso à internet.

No Brasil, existem muitas emissoras de rádio, mas há pouca cooperação entre elas para intercâmbio de material produzido. Uma parte significativa de programas produzidos pelas emissoras são muito valiosos para irem “ao ar” somente uma ou duas vezes. Estes produtos necessitam de maior divulgação de forma a atingir, formar e informar o maior número de pessoas. Uma das causas é a falta de um suporte técnico adequado para facilitar este intercâmbio. Com o advento da Internet, as redes de alta velocidade (banda larga) e principalmente com a possibilidade de compressão sem perda de qualidade dos arquivos de áudios, a transmissão de programas de longa duração (uma hora ou mais) deixou de ser um problema. O problema atual é como disponibilizar e preservar toda esta informação, possibilitando aos usuários um fácil acesso sem comprometer a segurança dos dados.

1.1 Objetivo Geral

Este estudo tem por objetivo analisar os problemas relacionados com o armazenamento digital a longo prazo dos arquivos de áudio e texto produzidos pela Rádio Senado e criar um exemplo de como estes arquivos juntamente com seus metadados podem ser preparados para o arquivamento, preservação e recuperação.

1.2 Objetivos Específicos

- Analisar os padrões existentes de codificação digital de áudio e sua aplicabilidade.
- Identificar métodos que possibilitem que a informação armazenada seja identificada, pesquisada e gerida de forma eficaz.
- Analisar os metadados utilizados na preservação de arquivos digitais e sua aplicação no modelo OAIS (Open Archive Information System). Este modelo tem se tornado um padrão a ser seguido visando a preservação a longo prazo de arquivos digitais, não se restringindo somente a áudio, mas a qualquer tipo de informação no formato digital.

- Elaborar um exemplo utilizando o padrão METS (Metadata Encoding & Transmission Standard) que possa ser utilizado como um pacote de informação a ser submetido ao um sistema OAIS.

1.2 Organização do trabalho

O capítulo 1 foi destinado à introdução, descrevendo os setores existentes na Rádio Senado, a situação atual e objetivos a alcançar.

No capítulo 2 abordaremos alguns padrões de formatos para áudios e textos digitais. Analisaremos os conceitos básicos de digitalização de áudio para facilitar a compreensão do tipo de arquivo ou informação que será armazenada, suas dimensões e espaço necessário para este armazenamento.

O capítulo 3 trata dos metadados necessários para a descrição, manutenção e recuperação a longo prazo de todas as informações armazenadas.

A segurança da informação é conceituada no capítulo 4, mostrando a necessidade de termos mecanismos de proteção dos dados.

No capítulo 5 estudaremos a preservação digital e as estratégias estruturais e operacionais visando a preparação do ambiente para o processo de preservação e as atividades e medidas concretas para a preservação.

O modelo de referência OAIS com seus pacotes de informação é apresentado no capítulo 6, com enfoque no processo relacionado com o produtor da informação, que é o primeiro passo para se proceder ao arquivamento das informações produzidas.

No capítulo 7 detalhamos um exemplo para um produto da Rádio Senado, composto por áudio e texto, que criamos com base no padrão METS para transmissão e codificação de metadados.

2 PADRÕES DE FORMATOS PARA OBJETOS DIGITAIS

2.1 FORMATOS PARA ÁUDIOS DIGITAIS

O áudio, assim como toda informação que é percebida pelos nossos sentidos, é analógico. O processo de capturar a informação analógica e transportá-la para o mundo digital é chamado de digitalização. Este processo deve considerar vários fatores de modo a preservar as características do áudio original, sendo que o principal fator é a nossa capacidade de audição, ou as frequências que nossos ouvidos conseguem captar, que estão na faixa de 20 a 20kHz.

A qualidade do áudio digitalizado é determinado basicamente por dois parâmetros:

1. taxa de amostragem: o número de vezes por segundo em que a amplitude da onda sonora é medida;
2. resolução: a quantidade de dígitos usados para armazenar cada medida.

A taxa de amostragem, descrita em kHz, deve ser no mínimo duas vezes o valor da maior frequência presente no sinal a ser digitalizado (Teorema de Nyquist). Como nosso ouvido percebe frequências de no máximo 20kHz, chegou-se a uma frequência mínima de amostragem de 44.1kHz (seria 40kHz, mas por limitações dos filtros dos circuitos eletrônicos chegou-se a este valor). Ou seja, serão gravados 44.100 valores numéricos por segundo.

A resolução define qual vai ser o maior valor numérico a ser utilizado na amostragem do sinal de áudio. Na linguagem dos computadores, significa a quantidade de bits a utilizar para armazenamento do dado digital. Quanto mais bits mais precisa será a medida a ser realizada. Utilizando-se 16 bits, temos 65.536 valores inteiros possíveis, resultando em uma numeração que varia de -32.768 a 32.767. Quanto maior a quantidade de bits maior a faixa dinâmica, menor ruído de fundo e maior fidelidade ao sinal original. Como uma posição de memória de um computador utiliza 8 bits (1 byte), duas posições de memória serão utilizadas para armazenar cada amostra, ou seja, 2 bytes.

A Figura 1 nos mostra o aumento da faixa dinâmica provocada pelo aumento da quantidade de bits de resolução e a Figura 2 ilustra processo de digitalização de um sinal analógico.

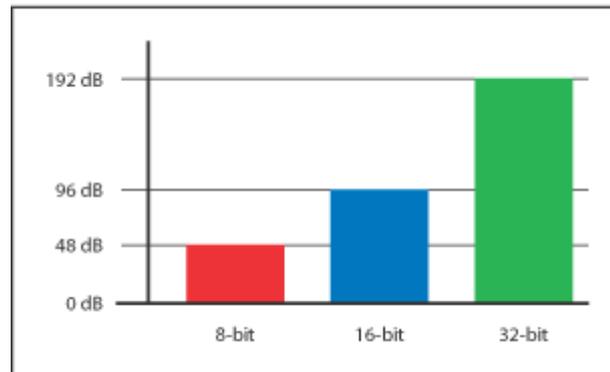


Figura 1 – Mais bits resulta em maior faixa dinâmica (fonte: ADOBE, 2003)

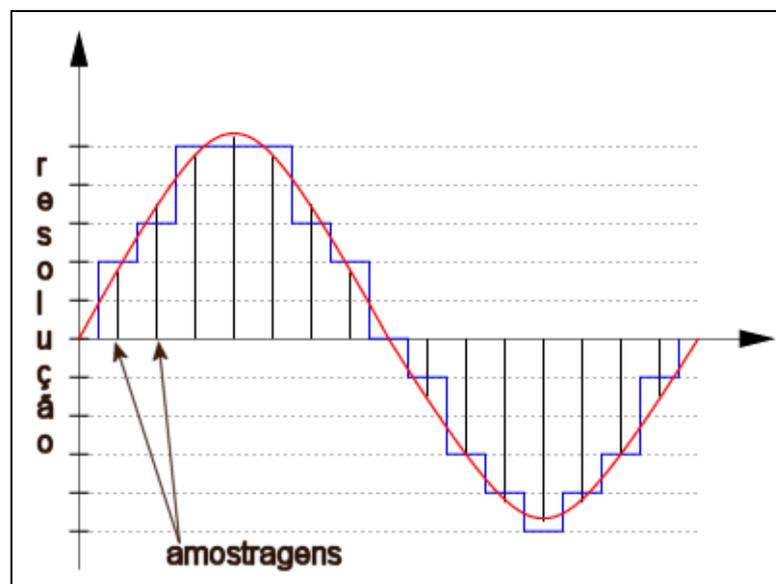


Figura 2 – Digitalização de um sinal de áudio

Um CD de áudio é produzido utilizando-se de uma taxa de amostragem de 44.1kHz e uma resolução de 16 bits, para cada um dos dois canais de áudio (esquerdo e direito), o que resulta em $44.100 \times 2 \times 2 = 176.400$ bytes por segundo, totalizando em um minuto 10.584.000 bytes (10MB) e em uma hora 600MB. Portanto, é uma quantidade de informação considerável a ser armazenada. Esta forma de armazenamento de áudio em formato digital é conhecida com PCM (Pulse Code Modulation) e não prevê o uso de nenhum tipo de compressão, gerando arquivos de grande tamanho. O ideal seria capturarmos o áudio utilizando altas taxas de amostragens e altas resoluções mas o seu armazenamento ficaria inviável economicamente como também dificultaria enormemente a disponibilização para transmissão ou *download* deste material.

Os formatos para áudio digital que abordaremos neste trabalho são os seguintes:

- a) WAVE (extensão wav) – formado por seqüência de valores numéricos que são interpretados pelo computador, possui alta qualidade e facilidade de edição, porém resulta em arquivos volumosos impróprios para transmissão via Internet. É o próprio arquivo PCM sem perdas ou compressão.
- b) MPEG-1 Layer III (extensão mp3) – são arquivos de áudio que sofrem compressão após a digitalização. Ocorrem perdas “controladas” na compressão, que consiste em retirar do áudio tudo o que o ouvido humano normalmente não perceberia. Os parâmetros de compressão do arquivo de áudio podem ser escolhidos pelo usuário de acordo com a aplicação e a qualidade desejada para o áudio. A taxa de compressão é medida em kbps (kilobits por segundo).

Para um sinal estéreo, amostrado a 44.1kHz, resolução de 16 bits, a taxa de bits de 128 kbps produz uma compressão de 11.0:1. Esta taxa é a mínima aceitável para uma emissora em FM e consegue reproduzir frequências até 16 kHz, condizente com a faixa de FM. Do mesmo modo, um sinal mono a uma taxa de 64kbps também sofre uma redução de 11.0:1.

A máxima qualidade em MP3 é conseguida a uma taxa de 320 kbps, que resulta em uma compressão de 4.4:1.

Em nosso exemplo utilizaremos o formato MP3 para armazenar os arquivos de áudio produzidos pela Rádio Senado. A Rádio Senado utiliza a taxa de 96kbps/mono para arquivos de áudio contendo somente voz tais como matérias, sonoras; 192 kbps/estéreo para os arquivos de áudio contendo programas, especiais e músicas.

2.2 FORMATOS PARA TEXTOS DIGITAIS

2.2.1 Arquivos TXT

Arquivos com a extensão TXT contém somente caracteres ASCII sem qualquer tipo de informação de formatação e é reconhecido por qualquer editor de texto e pode ser processado por qualquer tipo de software.

2.2.2 Arquivos DOC

Arquivos com a extensão DOC são usualmente criados por processadores de texto como Microsoft Word. Podem conter textos, imagens, fotos e qualquer outro tipo de informação comumente encontradas em editores de texto. O Microsoft Word utiliza formato proprietário para armazenar seus documentos. O armazenamento a longo prazo deste tipo de arquivo pode trazer problemas de compatibilidades com os novos softwares ou com novas versões do próprio pois no processo de conversão podem ocorrer perdas.

2.2.3 Arquivos PDF

Portable Document Format (PDF) é um formato de arquivo, desenvolvido pela Adobe Systems em 1993, para representar documentos de maneira independente do aplicativo, do *hardware* e do sistema operacional usados para criá-los. Um arquivo PDF pode descrever documentos que contenham texto, gráficos e imagens num formato independente de dispositivo e resolução. O PDF é um padrão aberto, e qualquer pessoa pode escrever aplicativos que leiam ou escrevam neste padrão. Há aplicativos gratuitos para Linux, Microsoft Windows e Apple Macintosh, alguns deles distribuídos pela própria Adobe (WIKIPEDIA, 2009).

O PDF/A (A de *Archive*) é uma especificação do formato PDF para armazenamento de longa duração, especificado pela norma ISO 19005-1-2005, que especifica o formato PDF como padrão universal para arquivamento de documento digital de longa duração.

O grande incentivador desta normatização foi o governo dos EUA, principalmente o judiciário e as agências reguladoras, que recebem diariamente uma enorme quantidade de páginas de documentos digitais em PDF e precisam garantir que estes documentos possam ser lidos, pesquisados e impressos daqui a 40, 50 anos ou mais.

Baseada na versão 1.4 do PDF da Adobe Systems, a norma nos diz o que pode e o que não pode estar em um PDF para que este esteja no padrão, eliminando dos documentos códigos de programação, elementos externos, fontes não desejadas etc.

Com o PDF/A, bibliotecas, centros de documentação, arquivos etc, poderão intercambiar seus conteúdos digitais com muita facilidade, pois o formato possui uma série de recursos que facilita esta tarefa. Entre estes recursos estão o suporte interno a metadados em XML, imagens supercomprimidas, imagens pesquisáveis pelo conteúdo e links de hipertexto (CAVALCANTI, 2006).

Em nosso exemplo utilizaremos o formato PDF/A para armazenar os arquivos de texto produzidos pela Rádio Senado.

2.3 FORMATOS PARA COMPRESSÃO SEM PERDAS

A compressão sem perdas permite gerar um arquivo de tamanho menor que o original para fins de transmissão ou armazenamento, possibilitando depois recriar o arquivo exatamente igual ao seu formato original. Como a maioria dos arquivos de computador possui a mesma informação registrada repetidas vezes, o programa de compressão lista esta informação uma vez, fazendo nova referência a ela sempre que voltar a aparecer no arquivo original, eliminando assim as redundâncias e reduzindo significativamente o tamanho do arquivo. Existem vários softwares que trabalham com este algoritmo de compactação tais como: PKZIP, WinRar, Winzip, 7-ZIP e BraZip (nacional).

Aqueles arquivos que incluem um conjunto de informações únicas, como gráficos ou arquivos de áudio, não podem ser muito comprimidos com este método porque não repetem muitos modelos de informação, entretanto para arquivos contendo texto estes algoritmos de compressão são muito eficientes.

Utilizaremos os arquivos no formato ZIP no nosso exemplo para gerar o Pacote de Submissão de Informação, que conterà os arquivos de texto, áudio e metadados a serem armazenados no sistema de preservação digital.

3 METADADOS

A indexação deve proporcionar a identificação de conceitos mais pertinentes ao conteúdo do documento produzindo uma correspondência precisa com o assunto pesquisado em índices. É a partir da realização desse processo que os resultados da questão de busca do usuário estarão condicionados. Ela é reconhecida como a parte mais importante do sistema porque condiciona os resultados de uma estratégia de busca. Além disso, a indexação pode ser observada em dois momentos distintos dentro do sistema: na entrada - no tratamento temático da informação - e na saída - na busca e recuperação da informação.

Para possibilitar a pesquisa de documentos multimídia foi necessário criar e definir padrões que permitissem descrever os conteúdos para que fosse possível que toda esta informação fosse identificada, pesquisada e gerida de uma forma mais eficaz.

O ato de pesquisar é essencialmente um processo que iguala os termos da pesquisa a palavras num documento. Então, caso os termos não sejam encontrados, nenhum resultado é apresentado mesmo que se relacione em parte com o termo pesquisado. Neste contexto existe um lugar reservado para o metadado.

Metadado pode ser entendido como a informação estruturada que descreve, explica, localiza, ou ainda possibilita que um recurso informacional seja fácil de recuperar, usar ou gerenciar. O termo metadado freqüentemente designa dados sobre dados, ou informação sobre informação. As funções dos metadados compreendem a descoberta de recursos - que permitem que recursos sejam identificados, localizados, selecionados por critérios de relevância e distinguidos por diferenças e similaridades; a organização de recursos; a facilitação da interoperabilidade; a identificação digital; e a preservação digital.

3.1 Tipos de metadados

Existem basicamente três tipos de metadados:

- Metadados descritivos: são os metadados que descrevem um recurso com o propósito de descoberta e identificação. Eles podem incluir elementos tais como título, autor, resumo e palavras-chave;
- Metadados estruturais: são os metadados que indicam como objetos compostos por vários elementos serão recompostos. Por exemplo, como as páginas de um livro, digitalizadas separadamente, são ordenadas para formar um capítulo;

- Metadados administrativos: fornecem informações que ajudam no gerenciamento de um recurso informacional. Por exemplo: informações sobre como e quando o recurso foi criado, informações técnicas sobre o arquivo e sobre quem possui direitos autorais sobre ele.

Existem vários subconjuntos de dados administrativos, mas dois deles têm sempre um destaque especial e são listados muitas vezes como tipos específicos de metadados, são eles: metadados para gerenciamento de direitos: fornecem informações sobre os direitos de propriedade intelectual relacionados a um determinado recurso; e metadados para preservação: guardam informações necessárias a gestão da preservação digital de um determinado recurso.

Metadados podem descrever recursos informacionais em qualquer nível de agregação – uma coleção, um recurso simples, ou um elemento que faz parte de um outro recurso, como, por exemplo, uma fotografia inserida num artigo; pode ainda ser usado para descrever um trabalho, bem como uma manifestação ou uma expressão desse trabalho, por exemplo, um relatório, uma edição particular desse relatório, ou uma cópia específica da edição. Os metadados podem estar embutidos num objeto digital – inscritos na sua codificação, como é comum nos documentos HTML (Hypertext Markup Language); ou podem estar armazenados separadamente, estruturados em base de dados e vinculados ao objeto que eles descrevem, facilitando a busca e a recuperação.

Gilliland-Swetkland (1998) dividiu os metadados em cinco categorias: Administrativo, Descritivo, Preservação, Técnico e Uso. O quadro da Tabela 1 apresenta as definições e exemplos para cada tipo.

Ainda segundo Gilliland-Swetkland (1998), os metadados possuem características diferentes, conforme exemplos mostrados no quadro da Tabela 2.

Tipo	Definição	Exemplos
Administrativo	Metadados usados no gerenciamento e administração de recursos de informação.	<ul style="list-style-type: none"> • Informação sobre aquisição. • Rastreamento da reprodução e dos direitos. • Documentação sobre requisitos de acesso legal. • Informação sobre localização. • Critérios de seleção para digitalização. • Controle de versões.
Descritivo	Metadados usados para descrever ou identificar recursos de informação.	<ul style="list-style-type: none"> • Registros de catalogação. • Guia de Arquivo.²⁴ • Índices especializados. • Relações de <i>hiperlinks</i> entre recursos. • Anotações feitas por usuários.
Preservação	Metadados usados no gerenciamento da preservação de recursos de informação.	<ul style="list-style-type: none"> • Documentação sobre a condição física dos recursos. • Documentação sobre ações tomadas para preservar versões físicas e digitais de recursos como, por exemplo, atualização e migração de dados.
Técnico	Metadados usados para retratar o funcionamento de um sistema ou comportamento dos metadados.	<ul style="list-style-type: none"> • Documentação de <i>hardware</i> e <i>software</i>. • Informação sobre digitalização, ex: formatos, taxas de compressão, rotinas de <i>scaling</i>. • Rastreamento dos tempos de resposta do sistema. • Autenticação e dados de segurança, ex: chaves de <i>encryption</i>.
Uso	Metadados usados para mapear o nível e tipo de uso dos recursos de informação.	<ul style="list-style-type: none"> • Registros de exibição. • Rastreamento do uso e de usuários. • Informação sobre múltiplas versões e reutilização de conteúdo.

Tabela 1 – Diferentes tipos de metadados e suas funções (fonte: Gilliland-Swetland, 1998)

Atributo	Características	Exemplos
Fontes	<i>Metadados internos</i> gerados por um agente criador para um objeto de informação no momento de sua criação ou digitalização.	<ul style="list-style-type: none"> • Nomes de arquivos e informação de cabeçalho. • Estruturas de diretório. • Formato de arquivo e esquema de compressão.
	<i>Metadados externos</i> relacionados a um objeto de informação, criados a <i>posteriori</i> , com frequência por alguém que não é o criador original.	<ul style="list-style-type: none"> • Registros de catalogação. • Direitos autorais e outras informações de cunho legal.
Método de criação	<i>Metadados automáticos</i> gerados por um computador.	<ul style="list-style-type: none"> • Índices de palavras-chaves. • <i>Logs</i> de transações do usuário.
	<i>Metadados manuais</i> criados por pessoas.	<ul style="list-style-type: none"> • Substitutos descritivos, tais como os registros de catalogação e os metadados Dublin Core.
Natureza	<i>Metadados não-profissionais</i> criados por pessoas que não são nem especialistas no assunto nem especialistas de informação, usualmente os criadores originais de um objeto de informação.	<ul style="list-style-type: none"> • <i>Metatags</i> criados para uma página <i>Web</i> pessoal. • Sistemas de arquivamento pessoais.
	<i>Metadados profissionais</i> criados ou por um especialista no assunto ou por especialistas de informação, usualmente não sendo o criador original do objeto de informação.	<ul style="list-style-type: none"> • Cabeçalho de assuntos especializados. • Registros MARC. • Guia de Arquivo.
Status	<i>Metadados estáticos</i> que nunca mudam a partir do momento em que foram criados.	<ul style="list-style-type: none"> • Título, proveniência e dados de criação de um recurso de informação.
	<i>Metadados dinâmicos</i> que podem mudar com o uso/manipulação de objetos de informação.	<ul style="list-style-type: none"> • Estrutura de um diretório. • <i>Logs</i> de transações de usuário. • Resolução de imagens.
	<i>Metadados de longa duração</i> necessários para assegurar que o objeto de informação continue a ser acessível e passível de utilização.	<ul style="list-style-type: none"> • Informação sobre processamento e formato técnico. • Informação sobre direitos autorais.
	<i>Metadados de curta duração</i> , especialmente de uma natureza transacional.	<ul style="list-style-type: none"> • Documentação referente ao gerenciamento da preservação.
Estrutura	<i>Metadados estruturados</i> que obedecem a uma estrutura previsível, padronizada ou não.	<ul style="list-style-type: none"> • MARC. • TEI e EAD. • Formatos de bases de dados locais.
	<i>Metadados não-estruturados</i> que obedecem a uma estrutura.	<ul style="list-style-type: none"> • Campos de notas e anotações não estruturadas.
Semântica	<i>Metadados controlados</i> que obedecem a um vocabulário padronizado ou a uma forma de autoridade.	<ul style="list-style-type: none"> • AAT. • ULAN. • AACR2.
	<i>Metadados não-controlados</i> que obedecem a um vocabulário padronizado ou a uma forma de autoridade.	<ul style="list-style-type: none"> • Notas de texto livres. • <i>Metatags</i> HTML.
Nível	<i>Metadados de coleção</i> relacionados às coleções de objetos de informação.	<ul style="list-style-type: none"> • Registro em nível de coleção, por exemplo, registro MARC ou guia de arquivo. • Índices especializados.
	<i>Metadados individuais</i> relacionados a objetos de informação individuais, usualmente contidos dentro de coleções.	<ul style="list-style-type: none"> • Legendas transcritas de imagens e datas. • Informação sobre formato.

Tabela 2 – Atributos e características dos metadados (fonte: Gilliland-Swetland, 1998)

3.2 Metadados existentes em arquivos de áudio

No caso dos arquivos de áudios, o formato MP3 tem embutido metadados chamados de *tag* ID3v1 e ID3v2 (Figura 3). Após a criação do padrão MP3 surgiu o problema de armazenar dados sobre o arquivo. O ID3 é um sistema de identificação que permite colocar informações relevantes sobre os arquivos de áudio dentro deles mesmos. O ID3v1 adicionou um pequeno bloco de 128 bytes ao fim do arquivo, mantendo a compatibilidade com os *players* existentes. Este bloco contém informações sobre título, intérprete, álbum e comentários, ano, gênero. O ID3v2 aperfeiçoou o ID3v1, tornando os campos de informações maiores e mais flexíveis e colocando estas informações no início do arquivo, possibilitando que elas fossem mostradas imediatamente quando transmitidas no modo *streaming*. Cada identificador ID3v2 contém um ou mais blocos de informações chamados quadros. Estes quadros podem conter qualquer tipo de informação e dados tais como título, álbum, intérprete, letra da música, imagens, etc.

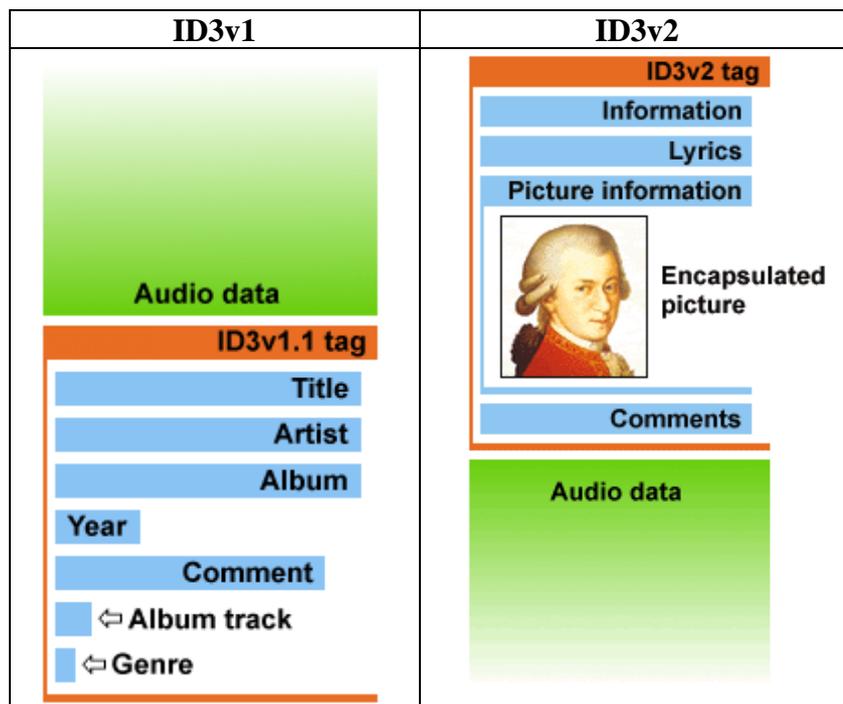


Figura 3 – Layout interno de um arquivo mp3 com ID3v1 e ID3v2 (fonte: <http://www.id3.org/ID3v2Easy>)

As Figuras 4 e 5 trazem como exemplo o software reprodutor de áudio WINAMP, mostrando os metadados na tela de reprodução e na tela de edição.

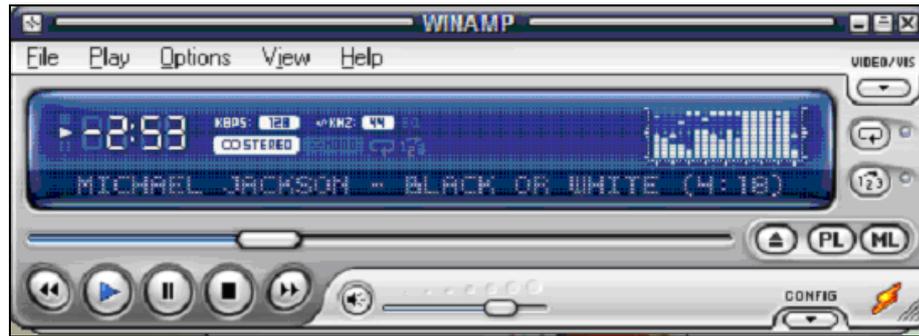


Figura 4 – Tela do software WINAMP mostrando metadados internos ao arquivo de áudio

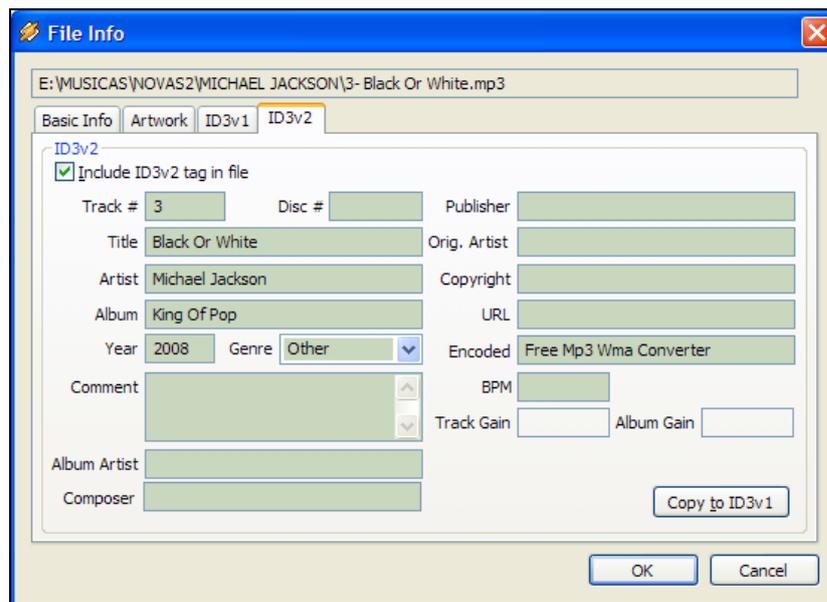


Figura 5 – Tela de edição dos metadados ID3V2

É essencial que algumas destas informações sejam preenchidas, tais como título, intérprete, compositores, ano, de forma que no momento em que estes arquivos forem solicitados e disseminados, não sejam perdidas as informações de sua origem e autoria. A maioria dos reprodutores de áudio (carro, internet, software de automação de rádios) utilizam-se das informações contidas nas *tags* ID3V1 e/ou ID3v2, principalmente título/intérprete para mostrar informações sobre a música que está em execução ou que foi executada. Os arquivos no formato WAV não possuem estes identificadores. Seu cabeçalho contém somente informações sobre os parâmetros necessários para a sua correta reprodução tais como taxa de amostragem, resolução, número de canais e tamanho do arquivo.

No entanto, para o armazenamento, preservação e recuperação dos arquivos de áudio é necessário que estas informações ou metadados estejam disponíveis nos Banco de Dados de modo a possibilitar o seu gerenciamento. Abordaremos, no próximo capítulo, a preservação digital com seus desafios e estratégias, bem como os requisitos necessários para a segurança das informações.

4 PRESERVAÇÃO DIGITAL

Da revisão da literatura pode-se extrair duas definições básicas de preservação digital ou arquivamento digital:

- Planejamento, alocação de recursos e aplicação de métodos de preservação e tecnologias necessárias para que a informação digital de valor contínuo permaneça acessível e utilizável por longo prazo, considerando-se neste caso longo prazo, o tempo suficiente para preocupar-se com os impactos de mudanças tecnológicas. A preservação digital aplica-se tanto a documentos "nato-digitais" quanto a documentos convertidos do formato convencional para o formato digital (HEDSTROM, 1998).
- Capacidade de manter a integridade e a acessibilidade da informação digital por longo prazo. Esta preservação da integridade e acessibilidade não se limita, apenas, a proteger a informação digital contra o acesso não autorizado mas, também, contra o uso inadequado resultante da má interpretação ou má representação da informação por parte dos sistemas computacionais. Percebe-se, aqui, o aspecto da inseparabilidade entre as atividades de preservação e acesso do mundo digital (TASK FORCE ON ARCHIVING OF DIGITAL INFORMATION, 1996).

Com a informatização acentuada de todas as atividades e conseqüentemente da forma como as informações são criadas e armazenadas, é grande o receio de que, dada a vulnerabilidade do ambiente digital, toda a informação gerada venha a ser perder ou se danificar ou corromper com o passar do tempo. Isto pode ocorrer devido a vários fatores:

- com a rápida evolução da tecnologia, a mídia onde está armazenada a informação torna-se obsoleta e não existem mais equipamentos compatíveis que façam a leitura dos dados armazenados;
- do mesmo modo, a informação pode estar em um formato não mais utilizado e não existem mais softwares compatíveis. Estes são problemas muito comuns e ocorreram com várias tecnologias anteriores (fitas de rolo, cassete, fitas DAT, minidisc) e posteriores à era da informática (disquetes, zip-drive, videodisco). O problema é que agora as mudanças estão ocorrendo em um espaço de tempo muito curto. Algumas tecnologias quando amadurecem já tem seu lugar ameaçado por outras melhores e de menor custo (HD-DVD x Blue Ray).

- A integração mundial através da internet tornou as bases de dados das instituições vulneráveis a ataques de 'hackers' que podem destruir todo um acervo digital que esteja desprovido da segurança adequada.

O acesso ao longo do tempo aos recursos digitais tem que ser assegurado pela existência de sistemas de hardware e software compatíveis com os mesmos. Da mesma forma, sem um conjunto de metadados documentando toda a informação existente torna-se praticamente impossível a recuperação destes recursos. A quantidade de informação gerada pelos sistemas informatizados é muito grande, e uma grande parte destas informações podemos considerar como sendo inútil ou 'lixo', o que cria a longo prazo uma série de dificuldades de armazenamento e de definição do que é necessário preservar. Na dúvida, armazena-se tudo, gerando conseqüentemente gastos desnecessários.

Um sistema de preservação digital para ser completo deve abranger todas as etapas pelas quais a informação percorre ao longo de sua existência. Desde a sua criação, passando pela manutenção, transformação e finalmente acesso pelo usuário final, os procedimentos adequados para a preservação da informação devem ser claramente definidos e seguidos.

As fases do ciclo de vida de um objeto de informação são (GILLILAND-SWETLAND, 1998):

- Criação de múltiplas versões: os objetos entram no sistema de informação digital criados na forma digital ou convertidos para a forma digital. Múltiplas versões do mesmo objeto podem ser criadas para preservação, pesquisa, disseminação ou até para desenvolvimento de produtos. Alguns metadados administrativos e descritivos podem ser incluídos pelo criador;
- Organização: os objetos são automaticamente criados ou manualmente organizados na estrutura de um sistema de informação digital e metadados adicionais podem ser criados através dos processos de registro, catalogação e indexação;
- Busca e recuperação: objetos armazenados e distribuídos são passíveis de busca e recuperação pelos usuários. Um sistema de computador cria metadados que rastreiam algoritmos de recuperação, transações de usuário e a eficácia dos sistema no armazenamento e recuperação;
- Utilização: objetos recuperados são utilizados, reproduzidos e modificados. Metadados referentes às anotações do usuário, mapeamentos dos direitos autorais e controle de versões podem ser criados;
- Preservação e disponibilização: os objetos de informação sofrem processos como revigoração, migração e checagem da integridade para assegurar sua contínua

disponibilidade. Objetos de informação que são inativos ou não mais necessários podem ser descartados. Os metadados podem documentar tanto as atividades de preservação quanto de disponibilização.

4.1 Estratégias da preservação digital

As estratégias propostas para enfrentar os desafios da preservação podem ser agrupadas em dois conjuntos: estratégias estruturais e operacionais. As estratégias estruturais dizem respeito aos investimentos ou esforços iniciais por parte da instituição, preparando seu ambiente para o processo da preservação digital. As estratégias operacionais, por sua vez, representam as atividades ou medidas concretas de preservação digital (THOMAS, 2004).

4.1.1 Estratégias estruturais

- Adoção de padrões: Esta estratégia recomenda a definição de um conjunto limitado de formatos para armazenar os dados; a utilização de padrões atuais para criar objetos digitais aliados à sua monitoração e migração para novos padrões já estabelecidos;
- Elaboração de manuais: estes manuais devem fornecer recomendações para as áreas de produção e arquivística, estabelecer as diretrizes para a seleção dos objetos a serem preservados, identificar as responsabilidades e descrever de forma completa os objetos digitais;
- adoção de metadados para preservação digital - facilita a pesquisa e a identificação das informação; manutenção e proteção dos dados; informações necessárias para representação do objeto; garantem a integridade da informação;
- Montagem de infra-estrutura para preservação digital: uma instituição será capaz de preservar seus objetos digitais através de uma infra-estrutura de *hardware*, *software* e pessoas que possua sistema de cópia de segurança, armazenamento redundante, detecção e recuperação automática de falhas; segurança de acesso físico e lógico e sistema hierárquico de armazenamento.

4.1.2 Estratégias operacionais

- Escolha do meio de armazenamento: as condições de acesso à informação irão determinar a disponibilidade (*online*¹, *near-line*², *off-line*³) além de uma série de outras características

¹ *online*: informações armazenadas em disco para acesso instantâneo, custo mais alto.

- físicas e lógicas devem ser avaliadas e ponderadas para se chegar a uma boa decisão, tais como: capacidade e forma física de armazenamento, durabilidade, robustez, tempo médio entre falhas, capacidade de detecção de falhas, capacidade de recuperação de falhas com interrupção/sem interrupção, presença de robótica, disponibilidade de suprimentos, assistência técnica, facilidade de manuseio, preço por unidade de armazenamento;
- Migração: é a estratégia mais utilizada para a preservação de objetos digitais. Consiste de um conjunto de atividades para copiar, converter ou transferir, periodicamente, a informação digital existente em uma determinada geração de tecnologia para as gerações subsequentes;
 - Emulação: refere-se à criação de novo *software* que imita o funcionamento do antigo *hardware* e/ou *software* para reproduzir seu comportamento;
 - Impressão em papel ou microfilme: esta estratégia poderia fixar o objeto como um todo, preservar o conteúdo e, de certa forma, o leiaute. A cópia digital melhora o acesso e a funcionalidade, enquanto a cópia em microfilme funciona como um substituto arquivístico;
 - Conservação da tecnologia: a manutenção da tecnologia, que criou os objetos, disponível para uso implicaria em requisitos de custo, espaço e suporte técnico impraticáveis. Este método seria uma medida transitória, enquanto não for possível a migração.

4.2 Segurança da Informação

A segurança da informação é a proteção dos sistemas de informação contra a negação de serviço a usuários autorizados, assim como contra a intrusão, e a modificação não-autorizada de dados ou informações, armazenados, em processamento ou em trânsito, abrangendo a segurança dos recursos humanos, da documentação e do material, das áreas e instalações das comunicações e computacional, assim como as destinadas a prevenir, detectar, deter e documentar eventuais ameaças a seu desenvolvimento (NBR 17999, 2003).

Estes são os princípios básicos para garantir a segurança das informações (NBR 17999, 2003):

- Confidencialidade – a informação somente pode ser acessada por pessoas explicitamente autorizadas;
- Disponibilidade – a informação ou sistema de computador deve estar disponível no momento em que a mesma for necessária;

² *near-line*: acesso quase em tempo real, usado para dados que podem ser recuperados em minutos.

³ *off-line*: acesso lento pois necessita da intervenção humana para sua operação.

- Integridade – a informação deve ser retornada em sua forma original no momento em que foi armazenada. É a proteção dos dados ou informações contra modificações intencionais ou acidentais não-autorizadas.

O item integridade não pode ser confundido com confiabilidade do conteúdo (seu significado) da informação. Uma informação pode ser imprecisa, mas deve permanecer íntegra (não sofrer alterações por pessoas não autorizadas). A segurança visa também aumentar a produtividade dos usuários através de um ambiente mais organizado, proporcionando maior controle sobre os recursos de informática, viabilizando até o uso de aplicações de missão crítica. A combinação em proporções apropriadas dos itens confidencialidade, disponibilidade e integridade facilitam o suporte para que as empresas alcancem os seus objetivos, pois seus sistemas de informação serão mais confiáveis.

Outros autores (DIAS, 2000; WADLOW, 2000; ALBUQUERQUE e RIBEIRO, 2002) defendem que, para uma informação ser considerada segura, o sistema que o administra ainda deve respeitar:

- Autenticidade – garante que a informação ou o usuário da mesma é autêntico; atesta com exatidão, a origem do dado ou informação;
- Não repúdio – não é possível negar (no sentido de dizer que não foi feito) uma operação ou serviço que modificou ou criou uma informação; não é possível negar o envio ou recepção de uma informação ou dado;
- Legalidade – garante a legalidade (jurídica) da informação; aderência de um sistema à legislação; característica das informações que possuem valor legal dentro de um processo de comunicação, onde todos os ativos estão de acordo com as cláusulas contratuais pactuadas ou a legislação política institucional, nacional ou internacional vigentes.
- Privacidade – foge do aspecto de confidencialidade, pois uma informação pode ser considerada confidencial, mas não privada. Uma informação privada deve ser vista / lida / alterada somente pelo seu dono. Garante ainda, que a informação não será disponibilizada para outras pessoas (neste caso é atribuído o caráter de confidencialidade à informação); é a capacidade de um usuário realizar ações em um sistema sem que seja identificado.
- Auditoria – rastreabilidade dos diversos passos que um negócio ou processo realizou ou que uma informação foi submetida, identificando os participantes, os locais e horários de cada etapa. Auditoria em software significa uma parte da aplicação, ou conjunto de funções do sistema, que viabiliza uma auditoria; consiste no exame do histórico dos

eventos dentro de um sistema para determinar quando e onde ocorreu uma violação de segurança.

Aliada à questão da segurança da informação, temos que considerar que o volume de informação em formato digital tende a aumentar consideravelmente com o tempo. Porém, a utilidade de toda esta informação depende do fato de ela poder ser filtrada e ser apresentada ao utilizador apenas a parte relevante da mesma.

Visando atender todos os requisitos necessários para a preservação digital a longo prazo, o modelo de referência OAIS (Open Archive Information System) foi criado para a NASA pelo CCSDS (Consultative Committee for Space Data Systems), para a preservação das informações digitais a respeito de clima, missões espaciais, pesquisas, documentos, tornando-se o padrão ISO 14721 em 2002. Este modelo é reconhecido como o mais importante trabalho conceitual de um sistema voltado para a preservação digital.

Descreveremos, em seguida, este modelo que encontra-se em grande expansão e que poderia ser aplicado com sucesso na preservação do acervo digital de áudio da Rádio Senado ou de quaisquer outros tipos de arquivos digitais.

5 MODELO OAIS (Open Archive Information System)

Um OAIS é um arquivo, consistindo de uma organização de pessoas e sistemas, que aceitou a responsabilidade para preservar a informação e torná-la disponível para uma determinada comunidade (CCSDS, 2002).

Este modelo de referência fornece uma estrutura para entendimento e consciência dos conceitos necessários para a preservação a longo prazo e acesso à informação digital. Possibilita que organizações não-arquivistas sejam participantes efetivos no processo de preservação e descreve e compara diferentes estratégias e técnicas de preservação. Ele aborda uma vasta gama de informações, incluindo funções de alimentação da informação (*ingest*), armazenamento (*archival storage*), gerenciamento de dados (*data management*), acesso (*access*) e disseminação (*dissemination*). Aborda também a migração da informação digital para novas mídias e formatos e modelos de dados usados para representar a informação.

O maior foco para a preservação das informações tem sido assegurar que elas estejam gravadas em mídias com longo prazo de estabilidade e que o acesso a estas mídias sejam cuidadosamente controladas.

5.1 O ambiente OAIS

O ambiente do modelo OAIS é composto de produtores, consumidores, administradores e o arquivo. Os produtores fornecem as informações a serem preservadas. Os consumidores solicitam os serviços do OAIS para buscar as informações de seu interesse. A administração estabelece a política do OAIS não se envolvendo com as operações de rotinas no arquivo, que são executadas por uma função dentro do próprio arquivo. A Figura 6 ilustra o ambiente em que o OAIS atua.

Uma clara definição da informação é essencial para a habilidade de um OAIS preservá-la. A informação é definida como qualquer tipo de conhecimento que pode ser trocado e esta informação é sempre representada por algum tipo de dado. Para entendermos o significado deste dado precisamos ter informações de como representá-los. Por exemplo, um CD-ROM pode ter vários dados gravados na forma de bits, mas para entendê-los precisamos saber como estes dados foram organizados, se na forma de imagem, áudio, vídeo, texto e qual o formato adotado.

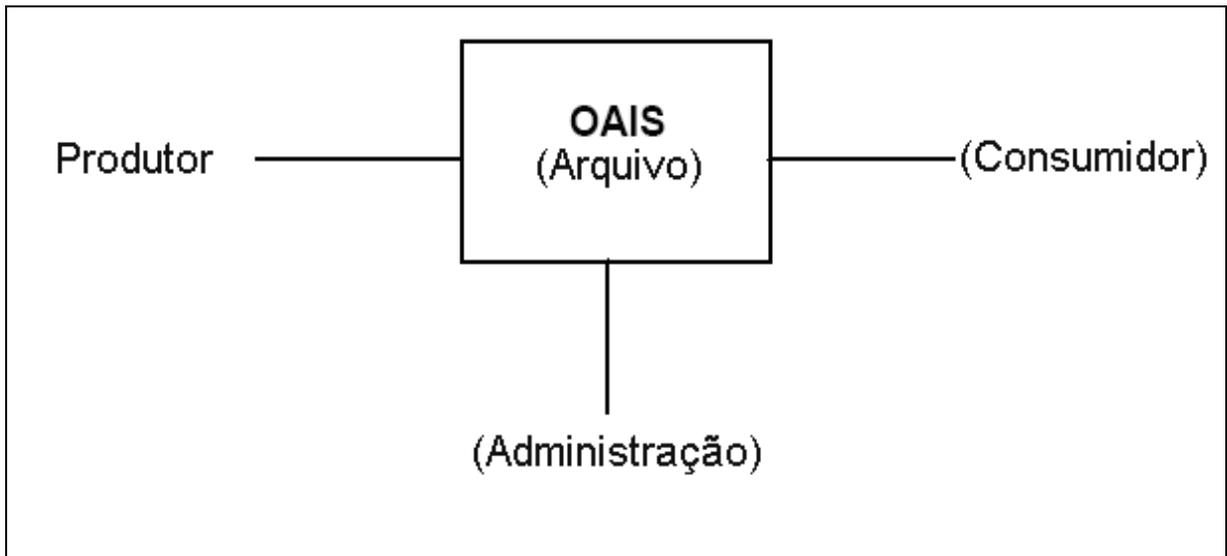


Figura 6 – O ambiente OAIS

Em geral, podemos dizer que “dados interpretados usando sua representação da informação produz informação”, conforme mostrado no esquema da Figura 7.

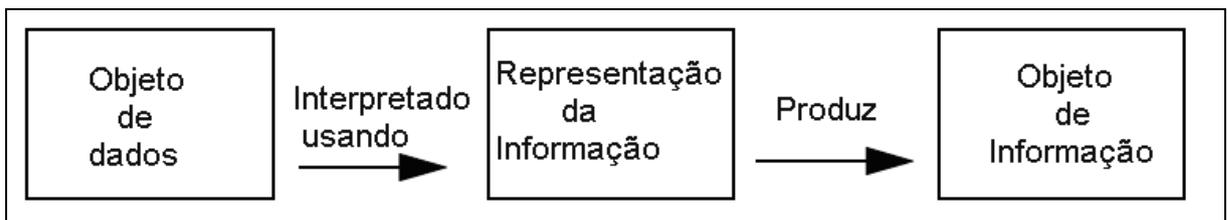


Figura 7 – Obtendo informações a partir dos dados armazenados

Para que o Objeto de Informação (Information Object) seja preservado com sucesso, é fundamental para um OAIS claramente identificar e entender o Objeto de Dado e sua Representação de Informação associada. Isto representa um desafio significativo para a preservação da informação digital.

5.2 Os Pacotes de Informação

Toda submissão de informação para o OAIS por um Produtor e toda disseminação de informação a um consumidor ocorre sob a forma de Pacotes de Informação.

Um Pacote de Informação é um contêiner conceitual de dois tipos de informação:

- Conteúdo da Informação(CI): informação sobre o objeto a ser preservado = objeto de dados mais a representação da informação;
- Informação de Descrição de Preservação (PDI): informação necessária para a preservação.

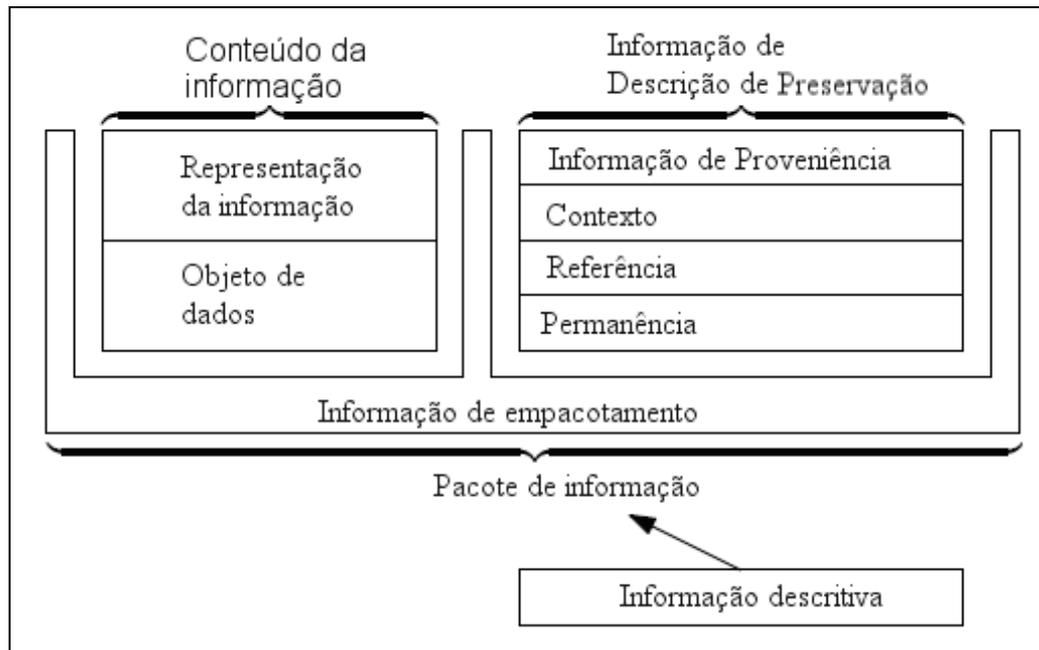


Figura 8 – Relacionamentos e conceitos do pacote de informação

A informação de empacotamento é aquela usada para unir, identificar e relacionar a CI e PDI. Exemplificando, se CI e PDI são identificados como sendo o conteúdo de arquivos em um CD-ROM, então a Informação de Empacotamento poderá incluir a estrutura de arquivos ISO9660 no CD-ROM. A Informação Descritiva é aquela usada para descobrir qual pacote tem a Informação de Conteúdo de interesse.

Os pacotes de informação no OAIS podem ser de três tipos:

- Pacote de Submissão da Informação (SIP⁴) a ser submetido a um OAIS;
- Pacote de Informação de Arquivo (AIP⁵): a ser preservado por um OAIS;
- Pacote de Disseminação da Informação (DIP⁶): a ser disseminado por um OAIS.

A Figura 9 mostra em mais detalhes as entidades funcionais do modelo OAIS e seus pacotes.

⁴ SIP - *Submission Information Package*

⁵ AIP - *Archival Information Package*

⁶ DIP - *Dissemination Information Package*

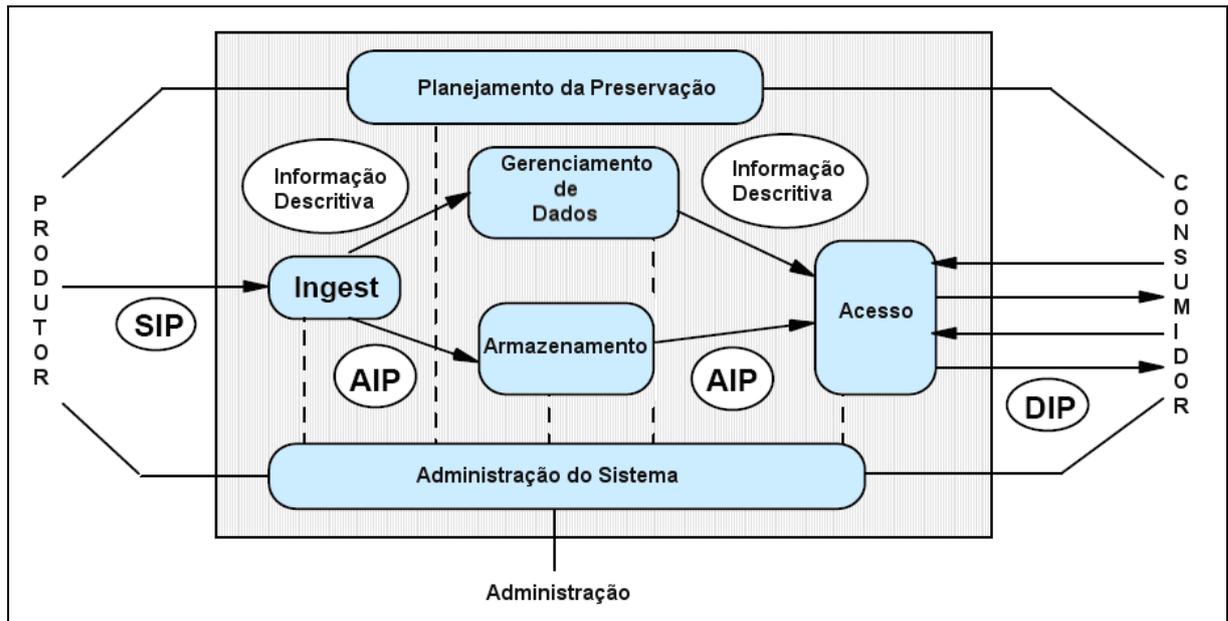


Figura 9 – Entidades funcionais do OAIS e os pacotes de informação (fonte: adaptado de CCSDS, 2002)

O Pacote de Submissão de Informação (SIP) é aquele que é enviado a um OAIS por um Produtor, sua forma e conteúdo são definidos entre o Produtor e o OAIS. Dentro do OAIS um ou mais SIPs são transformados em um ou mais AIPs para preservação. O AIP tem um conjunto completo de Informação de Descrição de Preservação (PDI) para a Informação de Conteúdo. Em resposta a uma solicitação de informação, o OAIS pode fornecer um ou vários AIP completos ou partes deles na forma de um Pacote de Disseminação de Informação (DIP).

Seis funções primárias são identificadas no modelo OAIS (CCSDS, 2002, p. 69):

- Recolhimento (*Ingest*) - faz a interface com o Produtor. Aceita pacotes SIP durante a sessão de submissão de dados pelo Produtor ou de componentes internos. Prepara os conteúdos para armazenamento e posterior acesso;
- Armazenamento (*Archival Storage*) – faz o armazenamento e manutenção dos pacotes AIP e fornece-os após solicitações.
- Gerenciamento dos Dados (*Data Management*) – aceita pacotes de descrição da função Ingest e outros tipos de metadados necessários para suportar as operações OAIS.
- Administração (*Administration*) – é a responsável pelo gerenciamento operacional do OAIS como um todo;
- Planejamento da Preservação (*Preservation Planning*) – responsável pelo monitoramento da evolução da tecnologia, criando estratégias e técnicas de garantam a preservação a longo prazo dos objetos digitais;
- Acesso (*Access*) – faz a interface com os Consumidores para identificação, localização e acesso às informações de interesse.

5.3 A entidade Produtor

Para que o Produtor submeta seus dados para armazenamento em um OAIS, antes deve se definir como a submissão será efetuada. Deverão ser definidos os conteúdos e formatos dos pacotes SIP que serão periodicamente transferidos nas sessões de submissão de dados. A sessão pode ser vista como conjuntos de objetos de dados e objetos descritivos. Em alguns casos a descrição ou metadados pode estar incluída no próprio objeto digital ou pode estar dividida em muitos itens descritivos separados.

Um padrão que tem se destacado na organização dos metadados necessários ao processo de preservação digital a longo prazo é o METS (*Metadata Encoding & Transmission Standard*). O METS define um padrão para a codificação de documentos XML que permite que seja empregado como qualquer um dos três tipos de pacotes de informação do modelo OAIS.

O padrão METS será apresentado com mais detalhes no próximo capítulo juntamente com um exemplo da codificação de um Pacote de Submissão de Informação para a Rádio Senado.

5.4 Exemplos de sistemas utilizando o modelo OAIS

aDORE

O sistema de repositório digital aDORE foi desenvolvido pelo *Research Library of the Los Alamos National Laboratory* para armazenar cópias digitais de publicações escolares entre outros tipos de trabalhos. Ele utiliza o método MPEG-21 DIDL de empacotamento de conteúdo.

DAITSS

O sistema de repositório digital DAITSS foi desenvolvido pela *Florida Center for Library Automation* como um suporte genérico para bibliotecas digitais ou repositórios institucionais. Ele utiliza o método METS para empacotamento do conteúdo e o esquema de metadados PREMIS para os metadados de preservação. Sua arquitetura é baseada no modelo de referência OAIS embora a função de acesso seja intencionalmente mínima.

DSpace

O sistema de repositório digital DSpace foi desenvolvido pela *MIT Libraries and Hewlett-Packard Laboratories* inicialmente para servir como um repositório para resultados de pesquisa. Utiliza o método METS para empacotamento do conteúdo. O uso do DSpace do modelo OAIS não está completo, encontra-se ainda está em fase de crescimento.

A Figura 10 mostra o diagrama funcional do DSpace.

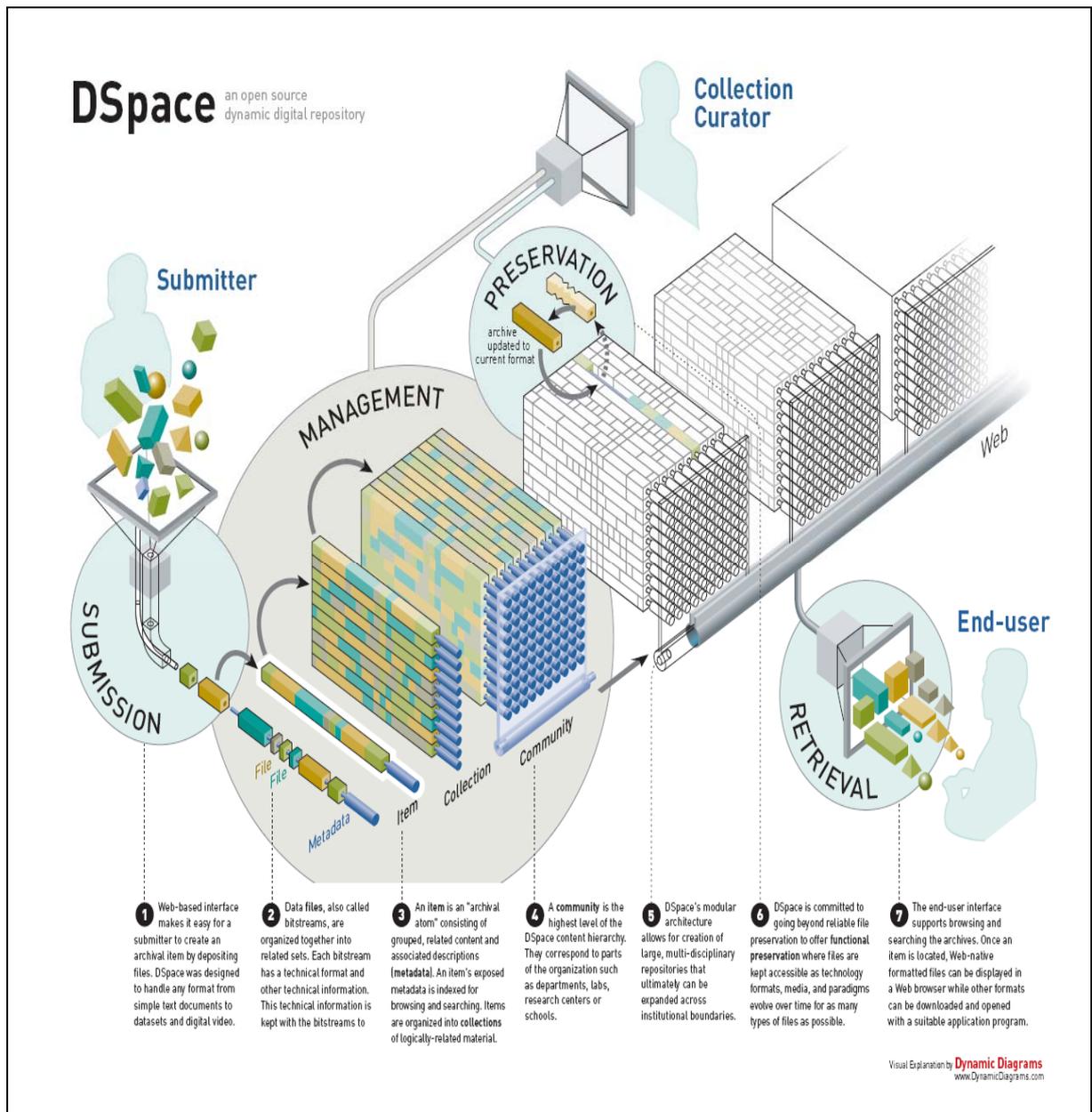


Figura 10 – Diagrama funcional do DSPACE (fonte: <http://www.dspace.org/images/stories/dspace-diagram.pdf>)

Fedora

O sistema de repositório digital Fedora foi desenvolvido pela *Cornell University* e a *University of Virginia*, inicialmente para armazenar coleções de bibliotecas digitais multimídias. Utiliza o método para empacotamento de conteúdo Fedora Object XML (FOXML), embora possa disseminar pacotes nos formatos METS e MPEG-21 DIDL.

A Figura 11 ilustra a estrutura interna do Fedora e suas interfaces com o Produtor, Consumidor e Administrador.

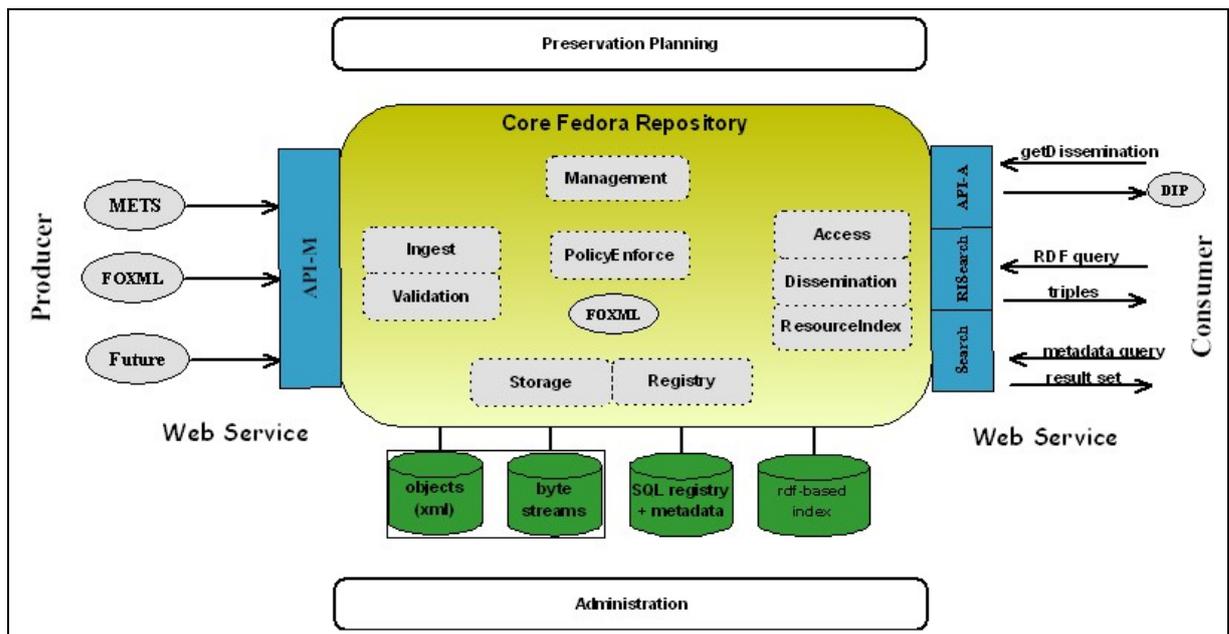


Figura 11 – Estrutura interna do Fedora (fonte: <http://fedora-commons.org/confluence/display/FCR30/Service+Framework>)

6 IMPLEMENTAÇÃO DE UM EXEMPLO

Como o modelo OAIS não especifica a forma de empacotar a informação, várias implementações têm sido propostas, tais como: METS, XFDU, MPEG-21 DIDL and IMS *Content Packaging*. Todas elas são baseadas na idéia de um arquivo “XML *manifest*” central que referencia ou contém os arquivos de dados que compõem o pacote.

Neste trabalho analisaremos uma forma de gerar um Pacote de Submissão de Informação (SIP), no formato METS, que seja compatível com o modelo OAIS e que tenha como objetos de dados, produtos da Rádio Senado tais como matérias jornalísticas, jornais e sonoras nos formatos de texto e áudio digitais.

Neste exemplo, usaremos como modelo o Jornal do Senado que é produzido para ser transmitido no horário destinado ao Senado Federal dentro do Programa “Voz do Brasil”, de segunda a sexta-feira, às 19h30 com duração de dez minutos. Este jornal é composto por um roteiro (arquivo de texto no formato doc) onde estão descritos todos os textos a serem lidos ‘ao vivo’ ou gravados e todas as matérias a serem veiculadas com seus respectivos áudios (formato mp3). O roteiro completo é mostrado no Anexo B, com destaque para os áudios de matérias⁷, sonoras⁸ e teaser⁹ utilizados no jornal.

Este exemplo servirá de modelo para os demais produtos da Rádio Senado.

6.1 METS

O *Metadata Encoding and Transmission Standard* (METS) foi desenvolvido pela *Digital Library Foundation* e é mantido pela Biblioteca do Congresso dos Estados Unidos.

O METS foi desenvolvido para prover uma estrutura de dados padronizada para descrever objetos digitais complexos no contexto de uma biblioteca digital. Este padrão especifica um formato em XML para codificar e encapsular os metadados descritivos, administrativos, e estruturais necessários tanto para a gestão de objetos de bibliotecas digitais como para a troca desses objetos entre repositórios ou entre repositórios e os seus utilizadores.

⁷ *Matérias são produtos jornalísticos em que o repórter entrevista, comenta ou descreve algum evento ocorrido no Senado, normalmente tem a fala de algum senador.*

⁸ *Sonoras são áudios extraídos de gravações de plenário, comissões ou entrevistas com o pronunciamento de algum senador.*

⁹ *Teaser é um recurso de rádio, televisão ou publicidade para estimular a curiosidade do público em relação a uma notícia, um programa, um anúncio, uma campanha, que só depois se farão conhecidos.*

Em um ambiente OAIS, um documento METS pode ser utilizado como um Pacote de Informação de Submissão, um Pacote de Informação de Arquivo ou um Pacote de Informação de Disseminação.

Um documento METS é composto pela raiz e de até sete seções que descrevem ou definem diferentes aspectos do objeto digital. A Figura 12 mostra a estrutura de um documento METS e a Figura 24 do Anexo A nos mostra algumas interligações possíveis dos elementos de um METS.

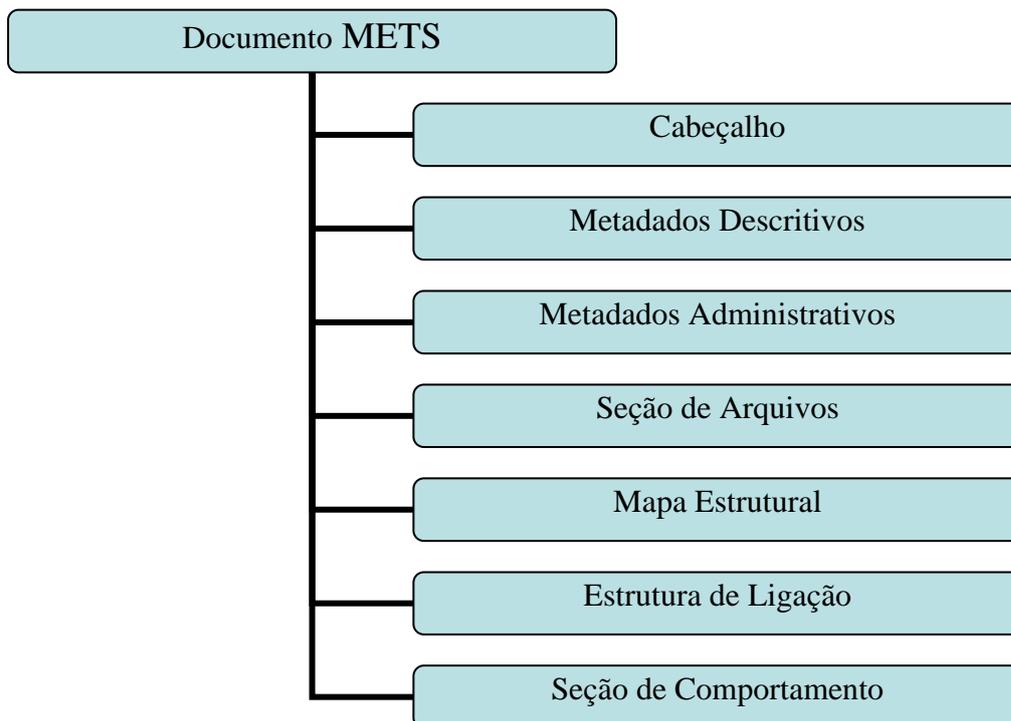


Figura 12 – Estrutura de um documento METS

Detalharemos cada seção do METS com base no exemplo que foi gerado para o jornal da Rádio Senado. Utilizamos o software XMLBlueprint¹⁰ (versão 7.1.0616) para fazer a criação, edição e validação de documentos no formato XML.

1 - Raiz (<mets>)

No topo do documento temos a indicação de que é um arquivo no formato XML, versão 1.0 e a codificação de caracteres (utf-8...) utilizada (Figura 13).

¹⁰ XMLBlueprint está disponível no endereço <http://www.xmlblueprint.com>.



```

1  <?xml version="1.0" encoding="utf-8"?>
2  [-] <METS:mets
3      xmlns="http://www.loc.gov/METS/"
4      xmlns:xlink="http://www.w3.org/1999/xlink"
5      xmlns:dc="http://purl.org/dc/elements/1.1/"
6      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
7      xmlns:METS="http://www.loc.gov/METS/"
8      xmlns:amd="http://www.loc.gov/AMD/"
9      xmlns:rights="http://cosimo.stanford.edu/sdr/metsrights/"
10     xsi:schemaLocation="
11         http://www.loc.gov/METS/
12         http://www.loc.gov/standards/mets/mets.xsd
13         http://www.eaudio.com/METS
14         http://www.eaudio.com/METS/mods-3-0.xsd
15         http://www.eaudio.com/METS
16         http://www.eaudio.com/METS/mix10.xsd
17         http://www.eaudio.com/METS
18         http://www.eaudio.com/METS/metsrights.xsd
19         http://www.eaudio.com/METS
20         http://www.eaudio.com/METS/AMD.xsd">

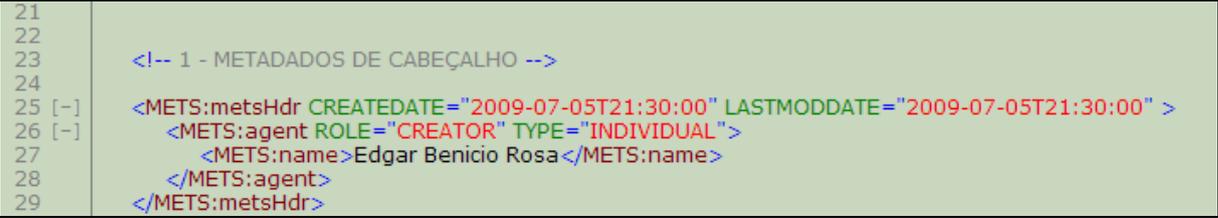
```

Figura 13 – Raiz de um documento METS

O XML provê uma forma eficiente de acessar padrões de metadados existentes por meio de *namespaces* (vocabulários controlados que identificam conceitos de maneira única, evitando ambiguidades). O *xmlns* é um atributo que associa um prefixo ao identificador – URL – de um *namespace*. Por exemplo, o prefixo “dc” se refere ao vocabulário do Dublin Core, fazendo referências a campos do Dublin Core (dc:title) que possam ser verificados sintaticamente pelo editor de XML.

2 - Cabeçalho METS (<metsHdr>) - contém a data em que o documento foi criado e atualizado, e detalha os agentes envolvidos na elaboração, tais como criador, editor, arquivista, disseminador. Não contém informação sobre o objeto a ser descrito.

A Figura 14 nos mostra o cabeçalho METS. As datas são escritas no formato “aaaa-mm-ddThh:mm:ss” indicando o dia e horário. O METS utiliza vocabulário controlado, assim os campos a serem preenchidos devem conter valores pré-definidos.



```

21
22
23  <!-- 1 - METADADOS DE CABEÇALHO -->
24
25 [-] <METS:metsHdr CREATEDATE="2009-07-05T21:30:00" LASTMODDATE="2009-07-05T21:30:00" >
26 [-]   <METS:agent ROLE="CREATOR" TYPE="INDIVIDUAL" >
27     <METS:name>Edgar Benicio Rosa</METS:name>
28   </METS:agent>
29 </METS:metsHdr>

```

Figura 14 – Metadados de cabeçalho

O campo ROLE especifica a função do agente a respeito do documento METS e os valores permitidos são descritos na Tabela 3.

Função	Descrição
CREATOR	as pessoas ou instituições responsáveis pelo documento.
EDITOR	as pessoas ou instituições que preparam os metadados para codificação.
ARCHIVIST	as pessoas ou instituições responsáveis pelo documento/coleção.
PRESERVATION	as pessoas ou instituições responsáveis pela preservação.
DISSEMINATOR	as pessoas ou instituições responsáveis pela disseminação.
CUSTODIAN	as pessoas ou instituições encarregadas da supervisão de um documento/coleção.
IPOWNER	as pessoas ou instituição que têm direitos sobre o objeto.
OTHER	usar OTHER se não for nenhuma das opções anteriores, pode ser definido no atributo OTHERROLE a definição adequada.

Tabela 3 – Agentes envolvidos no processo de preservação

O campo TYPE especifica o tipo de agente com os seguintes valores: INDIVIDUAL, ORGANIZATION, OTHER. O atributo opcional OTHERTYPE especifica o tipo de agente quando o valor OTHER é indicado no atributo TYPE.

3 - Seção de Metadados Descritivos (<dmdSec>) – descrevem as características e atributos do objeto digital. Vários padrões de esquemas de metadados podem ser utilizados, tais como: Dublin Core, MARC, MODS, TEI Header, EAD e outros.

A Figura 15 nos mostra a codificação dos metadados descritivos no formato METS.

O elemento *mdWrap* fornece um contêiner para metadados embutidos dentro de um documento METS, enquanto que o elemento *xmlData* é usado para armazenar conteúdos em formato XML.

Para os metadados de descrição optamos por utilizar o *Dublin Core* pois ele utiliza uma técnica de descrição simples e flexível, que pode ser facilmente estendida às mais complexas aplicações em vários domínios incluindo o ambiente audiovisual. Atualmente é o padrão mais conhecido e largamente usado para intercâmbio de metadados descritivos (JONG, 2003, p. 38). O SICON (Sistema de Informação do Congresso Nacional) utiliza o Dublin Core como opção de indexação de conteúdo numa ferramenta de busca integrada, possibilitando apresentar os resultados das pesquisas de uma forma homogênea, independente da heterogeneidade dos documentos pertencentes a bases de dados diferentes. O SICON é um conjunto de bases de dados produzidas no Senado Federal que permite a busca e consulta unificada ou em separado de documentos da rede RVBI (livros, artigos de revistas, recortes de jornais, multimídias, obras raras, etc.); discursos de senadores, legislação federal e matérias em tramitação no Senado.

```

30
31 <!-- 2 - METADADOS DESCRITIVOS - PADRÃO DUBLIN CORE (dc)-->
32
33 [-] <METS:dmdSec ID="EX_VB_01">
34 [-] <METS:mdWrap MDTYPE="DC" LABEL="Dublin Core Metadata">
35 [-] <METS:xmlData>
36 [-] <dc:dc>
37 <dc:title>Voz do Brasil</dc:title>
38 <dc:creator>Equipe da Voz do Brasil </dc:creator>
39 <dc:Subject>Voz do Brasil</dc:Subject>
40 <dc:Subject>Jornal do Senado</dc:Subject>
41 <dc:description>Roteiro e audio dos 10 minutos destinados ao Senado Federal no programa "Voz do
Brasil" </dc:description>
42 <dc:description>Arthur Virgilio anuncia que PSDB vai representar contra o Presidente José Sarney </dc:description>
43 <dc:description>Wellington Salgado diz que Presidente Sarney está sendo vítima de injustiça </dc:description>
44 <dc:description>Lider do governo exige que empresas e governos estaduais e municipais ajudem a União a estimular a
economia </dc:description>
45 <dc:publisher>Senado Federal </dc:publisher>
46 <dc:contributor> Radio Senado </dc:contributor>
47 <dc:date>2009-06-29</dc:date>
48 <dc:type>text; sound</dc:type>
49 <dc:format>pdf; mp3 </dc:format>
50 <dc:identifier>VB-20090629</dc:identifier>
51 <dc:source>gravacoes de audio de plenario e comissoes, materias e sonoras</dc:source>
52 <dc:language>pt </dc:language>
53 <dc:coverage>br </dc:coverage>
54 <dc:rights>Senado Federal </dc:rights>
55 </dc:dc>
56 </METS:xmlData>
57 </METS:mdWrap>
58 </METS:dmdSec>

```

Figura 15 – Metadados descritivos Dublin Core

O padrão Dublin Core é composto de 15 elementos básicos e diversos qualificadores. Cada elemento é opcional e pode ser repetido. A tabela 4 descreve as funções dos elementos deste padrão.

Elementos básicos	Descrição
1. Title	O nome pelo qual o recurso é formalmente conhecido
2. Subject	Assunto, expresso através de palavras-chaves ou códigos de classificação que descrevem o tópico do recurso. Recomenda-se o uso de vocabulário controlado.
3. Description	Descrição do conteúdo do recurso, podendo incluir um resumo, Tabela de conteúdo ou uma descrição livre.
4. Type	A natureza ou gênero do conteúdo do recurso: sound, text, image, etc. Recomenda-se o uso do vocabulário de tipos DCMI: http://dublincore.org/documents/dcmi-type-vocabulary/
5. Source	Uma referência ao recurso a partir do qual o presente recurso foi derivado, no todo ou em parte. Recomenda-se o uso de URI (Uniform Resource Identifier- http://www.ietf.org/rfc/rfc2396.txt)
6. Relation	Uma referência a um recurso relacionado. Recomenda-se uso de qualificadores tipo IsPartOf, IsVersionOf, Replaces, etc.
7. Coverage	A extensão ou escopo do conteúdo do recurso: localização espacial, período de tempo ou jurisdição.
8. Creator	A pessoa ou entidade responsável pela criação do conteúdo intelectual do recurso.
9. Publisher	Responsável pela disponibilização do recurso. Pode ser uma pessoa, uma organização ou um serviço.
10. Contributor	Responsável por contribuições ao conteúdo de um recurso. Pode ser uma pessoa, uma organização ou um serviço.
11. Rights	Informações sobre direitos autorais do recurso.
12. Date	Data de criação ou disponibilidade do recurso. Recomenda-se usar o formato YYYY-MM-DD (ano-mês-dia)

13. Format	Manifestação física ou digital do recurso. Pode ser o tipo de mídia ou as dimensões do recurso (tamanho ou duração). Pode ser usado para descrever o software, hardware ou outro equipamento necessário para visualizar ou operar o recurso. Exemplos: audio/mp3, audio/wav, application/pdf, text/plain.
14. Identifier	Referência única do recurso dentro de um dado contexto. Recomenda-se identificar o recurso por meio de letras ou números de acordo com um sistema de identificação formal, tais como URI, DOI, ISBN e ISSN.
15. Language	A linguagem utilizada no conteúdo do recurso. Usar código de 3 letras definido pela ISSO 639-2: http://www.loc.gov/standards/iso639-2/php/English_list.php

Tabela 4 – Elementos básicos no Dublin Core

4 - Seção de Metadados Administrativos (<amdSec>) – mantém informação sobre o gerenciamento dos objetos digitais possibilitando a utilização de quatro tipos de metadados: técnicos, direitos autorais, fonte e procedência digital. Metadados técnicos informam o formato do objeto (representação da informação no OAIS). Metadados fonte descrevem a fonte analógica de um objeto digital, enquanto a procedência cobre os relacionamentos entre arquivos dentro e externamente ao objeto, se o arquivo é derivado, manipulado ou transformado de outro. METS não descreve quais padrões de metadados devem ser usados aqui, ele somente fornece o contêiner para os metadados.

Para os metadados técnicos destinados a arquivos digitais de áudio utilizaremos o esquema denominado AUDIOMD, que tem por objetivo descrever arquivos de áudio digital e sua fonte digital ou analógica.

```

59
60 <!-- 3 - METADADOS ADMINISTRATIVOS - TÉCNICOS DE AUDIO - PADRÃO AUDIOMD (amd) -->
61 [-] <METS:amdSec>
62 [-]   <METS:techMD ID="FMT01">
63 [-]     <METS:mdWrap MDTYPE="OTHER" LABEL="FORMATO DO AUDIO">
64 [-]       <METS:xmlData>
65 [-]         <amd:AUDIOMD ANALOGDIGITALFLAG="FileDigital">
66           <amd:audio_data_encoding>mp3</amd:audio_data_encoding>
67           <amd:bits_per_sample>16</amd:bits_per_sample>
68           <amd:data_rate>96</amd:data_rate>
69           <amd:data_rate_mode>Fixed</amd:data_rate_mode>
70           <amd:duration>10:00</amd:duration>
71           <amd:num_channels>1</amd:num_channels>
72           <amd:sampling_frequency>44.1</amd:sampling_frequency>
73           <amd:sound_field>Monoaural</amd:sound_field>
74         </amd:AUDIOMD>
75       </METS:xmlData>
76     </METS:mdWrap>
77   </METS:techMD>
78

```

Figura 16 – Exemplo de metadados técnicos de um arquivo de áudio no formato MP3

A instituição responsável pelo esquema AUDIOMD é a Biblioteca do Congresso dos Estados Unidos e seus elementos estão descritos nas Tabelas 6 e 7. No exemplo da Figura 16, utilizamos somente os elementos suficientes para descrever um arquivo digital no formato MP3, mono, 96kbps, 16 bits por amostra, taxa de amostragem de 44.1kHz e com duração de 10 minutos.

Elementos	Descrição
audio_block_size	Tamanho de um bloco de áudio em bytes
audio_data_encoding	Estrutura dos dados de áudio; tipos conhecidos são PCM (pulse code modulation) e DSD ¹¹ (Direct Stream Digital) da SONY/PHILIPS.
bits_per_sample	Número de bits por amostragem do áudio: 16, 20, 24, etc.
calibration_ext_int	Indica se a informação de calibração está contida dentro do arquivo ou externamente.
calibration_location	Localização temporária do arquivo de calibração se ela for externa (URL)
calibration_type	Tipo de calibração, por exemplo, seqüência de testes ITU conhecidas como CCITT 0.33.00 (mono) e CCITT 0.33.01 (estéreo).
data_rate	Taxa de dados do áudio em um arquivo MP3 ou outro com compressão, expresso em kbps: 64, 128, 256, etc.
data_rate_mode	Indicador se a taxa de dados é fixa ou variável
duration	Duração do arquivo inteiro
first_sample_offset	Localização do primeiro byte válido de áudio no arquivo
first_valid_byte_block	Localização do primeiro byte válido de áudio no bloco
last_valid_byte_block	Localização do último byte válido de áudio no bloco
note	Informação adicional ou comentários sobre o arquivo de áudio
num_channels	Número de canais de áudio: 1, 2, 4, 5, etc. Ver elemento sound_channel_map.
sampling_frequency	Taxa na qual o áudio foi amostrado, expresso em kHz: 22, 44.1, 48, 96, etc.
sound_channel_map	Informação sobre a configuração dos canais, mapeando os canais de áudio de acordo com a posições dos alto-falantes: 1=left_front, 2=right_front, 3=center, 4=left_rear, 5=right_rear.
sound_field	Indica o arranjo espacial sonoro da gravação do áudio: monaural, stereo, joint stereo, surround sound DTS 5.1, etc.
word_size	Número de bytes em uma amostra do áudio, o qual geralmente é mapeada no elemento bits_per_sample.

Tabela 5 – Elementos do esquema de metadados para áudio (AUDIOMD)

¹¹ O formato de áudio DSD (Direct Stream Digital) é a base para o Super Audio CD (SACD), o formato para áudio de alta qualidade que é candidato a substituto dos CDs de música convencionais. Foi desenvolvido numa parceria entre a Sony e a Phillips.

O esquema RightsDeclarationMD contém metadados administrativos que descrevem as restrições/direitos de propriedade sobre um objeto digital. Os elementos principais são:

Elemento	Descrição
RightsDeclaration	descreve os direitos de propriedade associados com o recurso digital ou parte dele
RightsHolder	Pessoa ou organização proprietária de algum direito intelectual sobre o recurso
Context	Descrição das circunstâncias específicas associadas com as permissões e restrições

Tabela 6 – Esquema de metadados de direitos autorais (RightsDeclarationMD)

RightsDeclarationMD também contém três atributos: "RIGHTSDECID", "RIGHTSCATEGORY" e "OTHERCATEGORYTYPE". RIGHTSCATEGORY especifica a categoria do recurso: COPYRIGHTED, LICENSED, PUBLIC DOMAIN, CONTRACTUAL, ou OTHER.

A Figura 17 mostra um exemplo onde são utilizados alguns elementos deste metadado para descrever os direitos e informações de contato com o proprietário.

```

79
80 [-] <METS:rightsMD ID="RMD1">
81 [-] <METS:mdWrap MDTYPE="OTHER" LABEL="DIREITOS AUTORAIS">
82 [-] <METS:xmlData>
83
84 [-] <rights:RightsDeclarationMD RIGHTSCATEGORY="PUBLIC DOMAIN">
85 [-] <rights:RightsDeclaration>
86 Este objeto está disponível para uso público podendo ser reproduzido sem autorização.
87 </rights:RightsDeclaration>
88
89 [-] <rights:RightsHolder>
90 <rights:RightsHolderName>Senado Federal</rights:RightsHolderName>
91 [-] <rights:RightsHolderContact>
92 <rights:RightsHolderContactDesignation>Rádio Senado</rights:RightsHolderContactDesignation>
93 <rights:RightsHolderContactAddress>Praça dos Três Poderes. Anexo II - Bloco B - Brasília - DF.
CEP:70.165-900</rights:RightsHolderContactAddress>
94 <rights:RightsHolderContactEmail>radio@senado.gov.br</rights:RightsHolderContactEmail>
95 </rights:RightsHolderContact>
96 </rights:RightsHolder>
97
98 </rights:RightsDeclarationMD>
99
100 </METS:xmlData>
101 </METS:mdWrap>
102 </METS:rightsMD>
103
104 </METS:amdSec>

```

Figura 17 – Exemplo de metadados de direitos autorais

5 - Seção de Arquivos (<fileSec>) - esta seção relaciona todos os arquivos que compõem o objeto digital e onde podem ser encontrados (links).

No caso do Jornal do Senado, temos dois arquivos principais a serem relacionados: o roteiro e o arquivo de áudio. Observando o roteiro percebe-se a referência de outros arquivos de áudio que são utilizados durante o jornal. São áudios de entrevistas, pronuncionamentos de senadores, matérias gravadas em estúdios pelos repórteres, ou seja, são diversos áudios

produzidos pela emissora que podem ser veiculados isoladamente ou dentro de vários jornais produzidos naquele dia ou em outra data onde aquela informação for necessária. O produto final de um roteiro é sempre um arquivo único de áudio com a duração daquele jornal, fazendo parte dele todos os outros áudios. Então, neste exemplos consideramos importante colocar o áudio do jornal na íntegra, colocando também a referência dos demais áudios que foram utilizados para a geração deste produto. O roteiro foi convertido para o formato PDF/A e o arquivo de áudio está no formato MP3.

Foram criadas pastas numeradas sequencialmente no padrão Oyyyyymmdd-nnn para armazenamento dos arquivos VB-20090629.MP3, VB-20090629.PDF, MANIFEST.XML (este exemplo) e O20090629-001.ZIP, conforme mostrado nas Figuras 18 e 19. O arquivo O20090629-001.ZIP, contém os demais arquivos e é o pacote (SIP) a ser submetido em um sistema compatível com o modelo OAIS.

Pastas	Nome	Tamanho	Tipo	Data de modificação
O20090629-001	VB-20090629.mp3	4.495 KB	XMPlay-able file	29/6/2009 19:59
O20090629-002	VB-20090629.pdf	95 KB	Adobe Acrobat Document	6/7/2009 16:20
O20090629-003	manifest.xml	7 KB	Documento XML	20/7/2009 23:07
O20090629-004	O20090629-001.zip	4.532 KB	WinZip File	21/7/2009 10:16
O20090629-005				
O20090629-006				
O20090629-007				
O20090629-008				

Figura 18 – Pacote contendo áudio e texto

Pastas	Nome	Tama...	Tipo	Data de modificação
O20090629-001	20090629V09.mp3	110 KB	XMPlay-able file	29/6/2009 19:08
O20090629-002	20090629V09.xml	5 KB	Documento XML	8/7/2009 18:16
O20090629-003	O20090629-002.zip	111 KB	WinZip File	21/7/2009 10:29
O20090629-004				
O20090629-005				
O20090629-006				
O20090629-007				
O20090629-008				

Figura 19 – Pacote contendo somente áudio

O elemento <fileGrp> é utilizado para agrupar arquivos que são partes de um único arquivo, como no caso de grandes arquivos fracionados em vários menores. No exemplo da Figura 20, utilizamos dois grupos de arquivos, um para o texto (transcription) identificado como “ROTEIRO” e outro para os áudios (sound recording) identificado como “AUDIO”. Temos apenas um arquivo em cada grupo pois não temos fracionamento de arquivos.

```

105
106 <!-- 4 - SEÇÃO DE ARQUIVOS - PADRÃO AUDIO (AUDIOMD), TEXTO (TEI), VIDEO (VIDEO MD), IMAGEM (MIX) -->
107
108 [-] <METS:fileSec>
109
110 [-] <METS:fileGrp ID="ROTEIRO" USE="Transcription">
111 [-] <METS:file ID="FILE001" MIMETYPE="text/pdf" SIZE="257537" CREATED="2001-06-10T10:00:00">
112 <METS:Flocat LOCTYPE="URL" xlink:href="../O20090629-001/VB-20090629.pdf"></METS:Flocat>
113 </METS:file>
114 </METS:fileGrp>
115
116 [-] <METS:fileGrp ID="AUDIO" USE="sound recording">
117 [-] <METS:file ID="FILE002" MIMETYPE="audio/mpeg" SIZE="64232836" CREATED="2001-05-18T10:00:00" ADMID="FMT01">
118 <METS:Flocat LOCTYPE="URL" xlink:href="../O20090629-001/VB-20090629.mp3"></METS:Flocat>
119 </METS:file>
120 </METS:fileGrp>
121
122 </METS:fileSec>
123

```

Figura 20 – Exemplo de Seção de arquivos

O elemento <file> identifica os arquivos (ID), seu tipo (MIMETYPE), tamanho (SIZE), data e hora da criação (CREATED) e o subelemento <flocat> indica o tipo de endereçamento usado (LOCTYPE) e o local onde o arquivo se encontra armazenado (xlink:href) .

6 - Mapa Estrutural (<structMap>) – descreve a estrutura hierárquica do documento METS, fazendo referências aos arquivos informados na Seção de Arquivos. Permite fazer referências de diferentes versões da mesma seção lógica (por exemplo, arquivos de áudio e texto).

O mapa estrutural, mostrado nas Figuras 21 e 22, foi dividido nas seguintes partes: a divisão “div1” identifica o objeto como um todo com o nome “Voz do Brasil” que é subdividida em duas partes: a primeira parte composta pelas subdivisões “div2” e “div3” fazem referências ao roteiro (FILE001) e áudio (FILE002) já descritos na Seção de Arquivos; e a segunda parte composta pelas subdivisões “div4” e “div5”, mostradas na Figura 23, fazem referências às sonoras e matérias descritas em outro documento METS (elemento <mptr>) pois podem também ir “ao ar” isoladamente ou inseridas em outros jornais transmitidos pela emissora.

```

124
125 <!-- 5 - MAPA ESTRUTURAL -->
126
127 [-] <METS:structMap TYPE="logical">
128
129 [-] <METS:div ID="div1" LABEL="Voz do Brasil">
130
131 <!-- Jornal completo : roteiro + audio completos -->
132
133 [-] <METS:div ID="div2" LABEL="Voz do Brasil - roteiro" TYPE="text">
134 <METS:fptr FILEID="FILE001"></METS:fptr>
135 </METS:div>
136
137 [-] <METS:div ID="div3" LABEL="Voz do Brasil - audio completo" TYPE="sound recording">
138 <METS:fptr FILEID="FILE002"></METS:fptr>
139 </METS:div>
140

```

Figura 21 – Exemplo de Mapa Estrutural

```

141
142      <!-- Referências a outros documentos METS que descrevem as matérias e sonoras utilizadas neste Jornal -->
143
144 [-]      <METS:div ID="div4" LABEL="Sonoras" TYPE="sound recording">
145 [-]          <METS:mptr LOCTYPE="URL" xlink:href="../O20090629-002/O20090629-002.zip"
146              xlink:title="Teaser Sen. Arthur Virgilio"></METS:mptr>
147 [-]          <METS:mptr LOCTYPE="URL" xlink:href="../O20090629-003/O20090629-003.zip"
148              xlink:title="Teaser Sen. Wellington Salgado"></METS:mptr>
149      </METS:div>
150
151 [-]      <METS:div ID="div5" LABEL="Matérias" TYPE="sound recording">
152 [-]          <METS:mptr LOCTYPE="URL" xlink:href="../O20090629-004/O20090629-004.zip"
153              xlink:title="Sen. Arthur Virgilio e Wellington Salgado"></METS:mptr>
154 [-]          <METS:mptr LOCTYPE="URL" xlink:href="../O20090629-005/O20090629-005.zip"
155              xlink:title="Sen. Heráclito Fortes"></METS:mptr>
156 [-]          <METS:mptr LOCTYPE="URL" xlink:href="../O20090629-006/O20090629-006.zip"
157              xlink:title="Sen. Demóstenes Torres"></METS:mptr>
158 [-]          <METS:mptr LOCTYPE="URL" xlink:href="../O20090629-007/O20090629-007.zip"
159              xlink:title="Sen. Ideli"></METS:mptr>
160 [-]          <METS:mptr LOCTYPE="URL" xlink:href="../O20090629-008/O20090629-008.zip"
161              xlink:title="Sen. Paulo Paim"></METS:mptr>
162      </METS:div>
163
164  </METS:div>
165
166  </METS:structMap>
167

```

Figura 22 – Exemplo de Mapa Estrutural com referência a outros objetos METS

7 - Seção de Ligações Estruturais (<structLink>) – especifica “hyperlinks” entre as diferentes divisões de um mapa estrutural permitindo a representação das relações não hierárquicas existentes no objeto digital. Não se aplica ao exemplo em questão.

8 - Seção de Comportamento (<behavior>) – associa os arquivos com os códigos executáveis necessários para a sua leitura e manipulação. Corresponde à representação da informação do modelo OAIS. Não se aplica ao exemplo em questão.

O arquivo MANIFEST.XML completo é mostrado no Anexo C. Como mostrado anteriormente, este arquivo juntamente com o arquivo de roteiro VB-20090629.PDF e o de áudio VB-20090629.MP3 serão compactados gerando o arquivo O20090629-001.ZIP que deverá ser submetido a um sistema que utilize o modelo de referência OAIS, que se encarregará da preparação dos conteúdos para armazenamento e gerenciamento dentro do Arquivo. A Figura 23 ilustra o relacionamento do arquivo MANIFEST.XML com os demais arquivos descritos internamente e as referências existentes com outros pacotes de informação existentes.

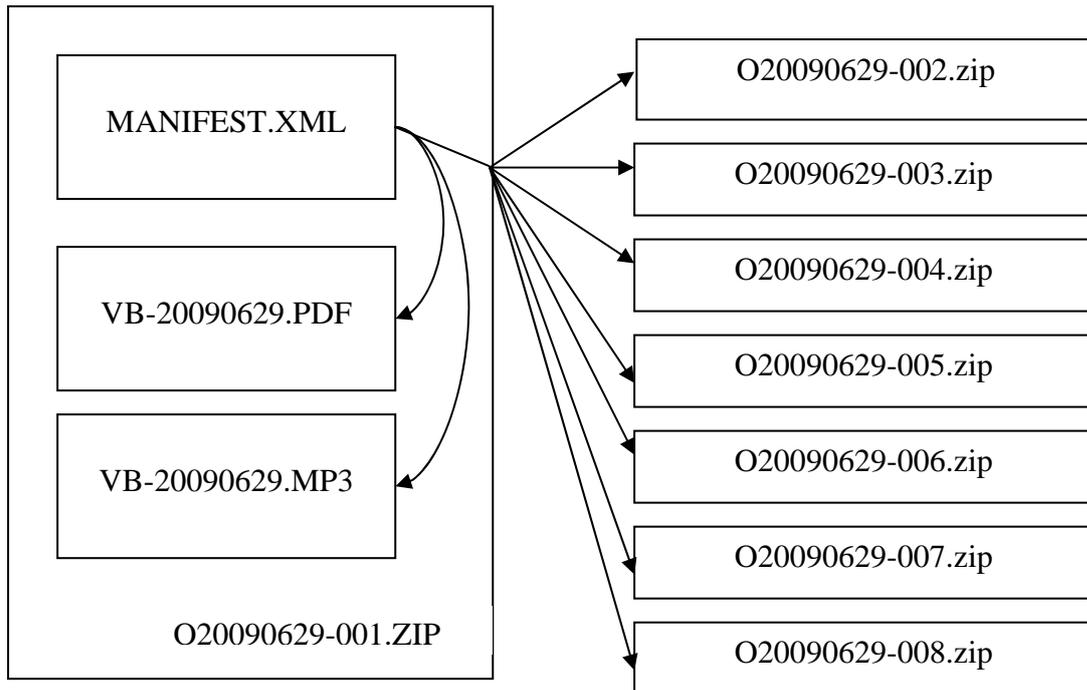


Figura 23 – Relacionamentos existentes no exemplo criado no padrão METS

Os arquivos METS para os demais áudios referenciados por este exemplo seguem o mesmo modelo aqui detalhado.

7 CONCLUSÃO

A preservação digital a longo prazo é um assunto vasto e complexo que tem obtido a especial atenção de várias entidades em todo mundo, preocupadas não somente em preservar o seu acervo como também em torná-lo útil e disponível a todos os interessados sem comprometer a segurança de seus dados.

O volume cada vez maior de informação em formato digital levará as instituições a adotarem estratégias para a preservação digital sob pena de possuírem grandes acervos mas estes serem subutilizados ou até mesmo irem se tornando inacessíveis ao longo do tempo. O ciclo de vida relativamente curto das novas tecnologias é uma ameaça real que assusta as entidades preocupadas com a preservação de seus acervos. Há a necessidade de se preservar o acesso ao objeto realizando o acompanhamento da evolução tecnológica.

Outras questões ligadas à preservação digital como as estratégias, metadados, proveniência, autenticidade e integridade de objetos digitais levaram a discussões do que seria mais apropriado para a preservação digital. O modelo OAIS surgiu para integrar os diversos esforços através de um modelo conceitual que servisse de base para o arquivamento de objetos digitais, possibilitando aumentar o grau de consciência e compreensão dos conceitos relevantes para o arquivamento, especialmente entre instituições não arquivísticas. Este modelo têm se apresentado como a proposta mais promissora de um padrão a ser seguido pelas grandes bibliotecas, institutos de pesquisa e várias outras entidades. Analisando o modelo OAIS, vimos que uma etapa muito importante em todo o processo é como a informação é criada e alimentada no sistema através dos pacotes de submissão de informação. Analisando o caso da Rádio Senado, criamos um exemplo utilizando o esquema METS para gerar um pacote contendo os dados a serem armazenados e os metadados necessários para a preservação. Este é um passo inicial dado a caminho de uma solução completa e ideal para o caso estudado.

7.1 Estudos Futuros

- Os elementos constantes dos metadados devem ser definidos juntamente com o pessoal responsável pelo arquivo de forma a conter todas as informações necessárias para a adequada preservação, manutenção e recuperação dos dados;
- Mecanismos de indexação devem ser previstos de forma a extrair dos objetos digitais a serem preservados, informações relevantes tais como: assunto, projeto de lei, normas, senadores, autoridades, possibilitando a recuperação e maximizando o aproveitamento deste

material. Como todo áudio tem associado um texto, esta indexação é simples de ser realizada;

- As demais entidades funcionais do modelo OAIS devem ser estudadas e analisadas por todos os setores envolvidos no processo de preservação de forma a adquirir um conhecimento mais aprofundado deste modelo;
- Estudar, realizar testes nos aplicativos DSpace, Fedora e em sistemas similares que utilizam o modelo OAIS, possibilitando propor a solução ideal e, principalmente, possível de ser implementada para a Rádio Senado;
- Analisar a possibilidade do sistema OAIS exportar as informações armazenadas sobre os programas e matérias (no formato Dublin Core) para o sistema SICON - Sistema de Informação do Congresso Nacional, proporcionando a integração com outros sistemas existentes e uma maior divulgação das informações armazenadas.

REFERÊNCIAS BIBLIOGRÁFICAS

ADOBE. **Adobe Audition 1.0: A Digital Audio Primer**. 2003. Disponível em: <<http://www.adobe.com/products/audition/pdfs/audaudioprimer.pdf>> Acesso em: 29 jul. 2009.

ALBUQUERQUE, Ricardo e RIBEIRO, Bruno. **Segurança no Desenvolvimento de Software – Como desenvolver sistemas seguros e avaliar a segurança de aplicações desenvolvidas com base na ISO 15.408**. Editora Campus. Rio de Janeiro, 2002.

CAVALCANTI, Luiz Augusto Bellucci. **PDF/A - O padrão ISO para Arquivamento de longo prazo**. 2006. Disponível em: <<http://www.pdf.com.br/?pag=99&sub=101>>. Acesso em: 12 jul. 2009.

CCSDS, Consultative Committee for Space Data Systems. **Reference Model for an Open Archival Information System (OAIS)**. Blue Book. 2002. Disponível em: <<http://public.ccsds.org/publications/archive/650x0b1.pdf>>. Acesso em: 22 jul. 2009.

DIAS, Cláudia. **Segurança e Auditoria da Tecnologia da Informação**. Axcel Books. Rio de Janeiro, 2000.

FLEISCHHAUER, Carl. **Digital Formats for Content Reproduction**. Washington, DC: Library of Congress, 1998. Disponível em <<http://lcweb2.loc.gov/ammem/formats.html>>. Acesso em: 04 abr. 2008.

GILLILAND-SWETLAND, Anne J. **Defining Metadata. In: Introduction to Metadata: Pathways to Digital Information**. Califórnia, 1998.

HEDSTROM, Margaret. **Digital preservation: a time bomb for digital libraries**. 1998. Disponível em: <<http://www.uky.edu/~kiernan/DL/hedstrom.html>>. Acesso em: 15 jul. 2009.

JONG, Annemieke de. **Metadata in the audiovisual production enviroment. An Introduction**. Nederlands Instituut voor Beeld en Geluid. 2003. Disponível em: <http://archivesatrisk.org/restricted/standards/Metadata_En_version_2003.pdf>. Acesso em: 22 jul. 2009.

NBR ISO/IEC 17799 – **Tecnologia da Informação. Código de Prática para Gestão da Segurança da Informação**. Associação Brasileira de Normas Técnicas. Rio de Janeiro, 2003.

NBR ISO/IEC 17799 – **Tecnologia da Informação. Código de Prática para Gestão da Segurança da Informação**. Associação Brasileira de Normas Técnicas. Rio de Janeiro, 2003.

NOERR, Peter. **The Digital Library Toolkit**. Sun Microsystems. 2 rd edition. Santa Clara, CA.: Sun Microsystems, March 2000. Disponível em <http://daminfo.wgbh.org/digital_library_toolkit.pdf>. Acesso em: 04 abr. 2008.

RODRIGUES, Nelson de Almeida - **INTRODUÇÃO AO METS – Preservação e Intercâmbio de Objetos Digitais**. Brasília, 2008. Disponível em: <<http://www.periodicos.ufsc.br/index.php/eb/article/viewFile/1885/6644>>. Acesso em: 13 jul. 2009.

SANTIAGO, Mônica Cristina Costa. **Metadados para Recuperação da Informação em Ambiente Virtual**. Rio de Janeiro, 2004. Disponível em: <http://tede-dep.ibict.br/tde_busca/arquivo.php?codArquivo=25>. Acesso em: 13 jul. 2009.

TASK FORCE ON THE ARCHIVING OF DIGITAL INFORMATION. *Preserving digital information; report of the Task Force on Archiving of Digital Information*. Washington, D.C, 1996. Disponível em: <<http://www.clir.org/pubs/reports/pub63watersgarrett.pdf>>. Acesso em: 22 jul. 2009.

THOMAZ, Katia P; SOARES, Antonio José. **A preservação digital e o modelo de referência Open Archival Information System (OAIS)**. DataGramaZero - Revista de Ciência da Informação - v.5 n.1 fev/2004 - ARTIGO 01. Disponível em: <http://dgz.org.br/fev04/Art_01.htm>. Acesso em: 12 jul. 2009.

WADLOW, Thomas. **Segurança de Redes**. Editora Campus. Rio de Janeiro, 2000.

WIKIPEDIA. **Portable Document Format**. Disponível em: <http://pt.wikipedia.org/wiki/Portable_Document_Format>. Acesso em: 12 jul. 2009

ANEXO A

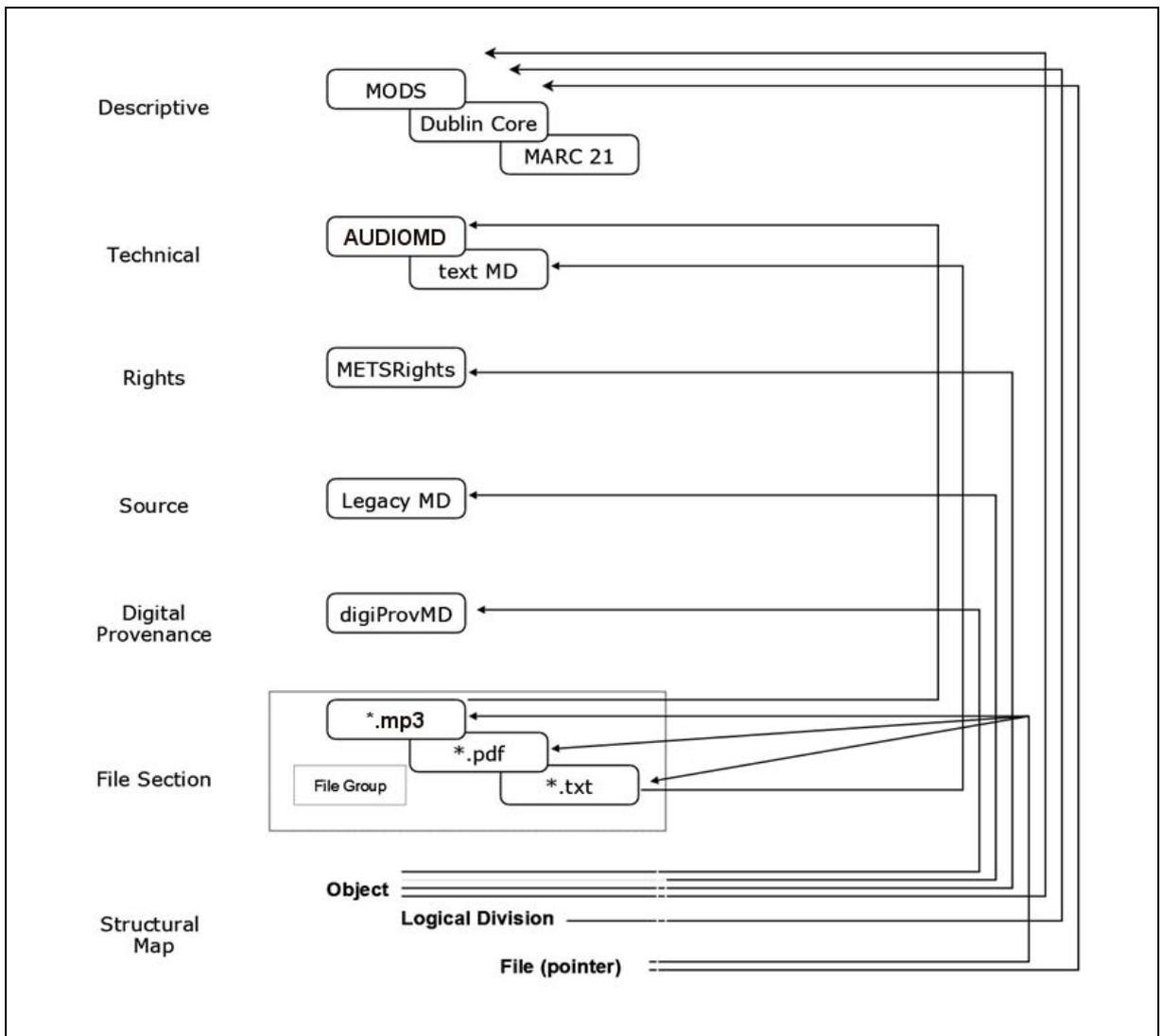


Figura 24 – Ligações existentes em um documento METS (fonte: www.jisc.ac.uk)

ANEXO B

ROTEIRO PARA O JORNAL DO SENADO NO PROGRAMA “VOZ DO BRASIL”

29/06/2009

SEGUNDA-FEIRA

LOC¹²: JORNAL DO SENADO

LOC: ARTHUR VIRGÍLIO ANUNCIA QUE PSDB VAI REPRESENTAR CONTRA O PRESIDENTE JOSÉ SARNEY:

Comentário 01 = áudio de curta duração com fala de senador¹³

TEASER: (0629V09 - ARTHUR VIRGÍLIO): “QUERO INVESTIGAÇÃO DURA SOBRE TODOS OS ATOS E AS CORRELAÇÕES POSSÍVEIS DE TODOS OS PRESIDENTES DA CASA QUE PASSARAM PELA GESTÃO DO SR. AGACIEL MAIA, DE 14 ANOS.”

LOC: WELLINGTON SALGADO DIZ QUE PRESIDENTE SARNEY ESTÁ SENDO VÍTIMA DE INJUSTIÇA:

Comentário 02 = áudio de curta duração com fala de senador

TEASER:” (0629V10 - WELLINGTON SALGADO): “FOI INJUSTO COM O PRESIDENTE SARNEY. O PRESIDENTE TÁ PAGANDO UMA CONTA QUE NÃO É DELE.”

LOC: LÍDER DO GOVERNO EXIGE QUE EMPRESAS E GOVERNOS ESTADUAIS E MUNICIPAIS AJUDEM UNIÃO A ESTIMULAR ECONOMIA

TÉC¹⁴: VINHETA

¹² LOC: indica os textos a serem lidos pelos locutores

¹³ Teaser: s.m. (pal. ing.) Recurso de rádio, televisão ou publicidade para estimular a curiosidade do público em relação a uma notícia, um programa, um anúncio, uma campanha, que só depois se farão conhecidos.

¹⁴ TEC: indica qual o áudio o operador deve reproduzir.

BOA NOITE!! O LÍDER DO PSDB, SENADOR ARTHUR VIRGÍLIO, ANUNCIOU QUE VAI REPRESENTAR NO CONSELHO DE ÉTICA DO SENADO CONTRA O PRESIDENTE DA CASA, JOSÉ SARNEY, POR ACREDITAR QUE ELE TEM SIDO OMISSO EM RELAÇÃO ÀS DENÚNCIAS DE IRREGULARIDADES NA CASA.

ARTHUR VIRGÍLIO EXIGIU AINDA O IMEDIATO AFASTAMENTO DE SARNEY DA PRESIDÊNCIA DO SENADO ATÉ QUE TODAS AS DENÚNCIAS SEJAM ESCLARECIDAS, COMO INFORMA O REPÓRTER ALEXANDRE CAMPOS.

Comentário 03 = áudio de matéria realizada pelo repórter Alexandre

TÉC: (0629A19 – Virgílio / Salgado – Alexandre – 1’47”) o líder do psdb acredita ainda que existe uma ação para intimidá-lo e desmoralizá-lo, com denúncias infundadas contra ele, e suspeita de que há senadores envolvidos no esquema com o ex diretor-geral da casa, **agaciel maia. (virgílio)** *nós temos o dever de saber quais são os senadores covardes, corruptos que protegeram esse desmando o tempo inteiro, porque não ficou aí sozinho. Porque jabuti não sobe em árvore. Porque jabuti quando está em árvore é enchente ou mão de gente. Então tem senador sim que apadrinhou esse corrupto para fazer um roubo que foi de usufruto não de apenas dele, deve ter dividido com muita gente com assento com mandato nesta casa. (repórter) arthur virgílio apresentou outras sugestões à mesa diretora para moralizar o senado, como o cadastramento de todos os servidores efetivos, comissionados e terceirizados, bem como a redução de cargos à disposição dos senadores. o senador pediu também a investigação de todos os senadores que ocuparam o cargo de primeiro-secretário quando agaciel maia era diretor-geral do senado. (virgílio 2) quero uma investigação dura sobre todos os Primeiros Secretários que passaram por essas gestões junto com Agaciel Maia. Quero investigação dura sobre todos os atos e as correlações possíveis de todos os Presidentes da Casa que passaram pela gestão do Sr. Agaciel Maia, de 14 anos. Quero a saída da Presidência da Casa do Presidente José Sarney. (repórter) já o senador wellington salgado, do pmdb mineiro, lembrou a arthur virgílio que o presidente José Sarney tem tomado todas as providências para apurar as denúncias e acrescentou que as supostas irregularidades foram praticadas em gestões anteriores, sem qualquer ligação com a atual mesa diretora. (salgado) senador arthur virgílio o discurso de vossa excelência é de honra, mas foi de injustiça. foi injusto com o presidente sarney. o presidente tá pagando uma conta que não é dele. é anterior, vossa excelências tem que denunciar brigar. mas estão sendo injusto com o presidente sarney. ele tem feito tudo o que os senadores tem pedido. tudo. vossa excelência sabe disso. tira fulano, bota sicrano. muda, agora tem que aprovar*

diretor.

LOC: O PRIMEIRO SECRETÁRIO DO SENADO, HERÁCLITO FORTES, EM APARTE, GARANTIU A ARTHUR VIRGÍLIO QUE TODAS AS PROVIDÊNCIAS QUE ELE EXIGE JÁ FORAM TOMADAS E ESTÃO EM FASE DE APURAÇÃO.

LOC: O PRESIDENTE DO SENADO, JOSÉ SARNEY, VAI COBRAR DOS LÍDERES PARTIDÁRIOS QUE INDIQUEM OS SENADORES PARA O CONSELHO DE ÉTICA. O CONSELHO ESTÁ PARALISADO DESDE MARÇO, QUANDO VENCEU O MANDATO DE DOIS DE SEUS MEMBROS.

O ANÚNCIO FOI FEITO NA MANHÃ DESTA SEGUNDA-FEIRA PELO PRIMEIRO-SECRETÁRIO DA CASA, HERÁCLITO FORTES, COMO EXPLICA O REPÓRTER GEORGE CARDIM.

Comentário 04 = áudio de matéria realizada pelo repórter Cardim

TÉC: (0629A17 CARDIM HERÁCLITO CONSELHO DE ÉTICA TEMPO:) **(CORTA)** o primeiro-secretário, senador heráclito fortes, do democratas do piauí, disse que o presidente da casa, josé sarney, deve voltar a cobrar dos líderes dos partidos que indiquem os 15 titulares e 15 suplentes ainda nesta semana **(heráclito)** ***“eu tenho impressão que esta semana o presidente da casa cobre estas indicações para que o conselho de ética volte a funcionar o mais rápido possível”***. (cardim) (corta) o primeiro-secretário anunciou também que deve analisar e divulgar nesta semana os resultados da comissão de sindicância criada para investigar os chamados atos secretos no senado. **(CORTA)** ao contrário da recomendação da procuradoria geral da república, que defende a anulação de todos os atos não publicados, heráclito esclareceu que muitos não podem ser cancelados, pois já perderam até a validade. no entanto, o senado vai buscar identificar e punir os responsáveis por atos ilegais, que prejudicaram o funcionamento ou causaram prejuízos à casa. **(heráclito)** ***eu quero ponderar que alguns destes atos deixaram despesas geraram malefícios ao erário público e alguém precisa se responsabilizar por esta anulação. nós temos um caso concreto de um uma denúncia feita pelo senador demóstenes do uso indevido de uma função de seu gabinete. isto é crime, precisa ser apurado e os culpados punidos”*** (CORTA).

LOC: EM CARTA AOS SENADORES, O PRESIDENTE JOSÉ SARNEY INFORMA QUE PEDIU À POLÍCIA FEDERAL QUE INVESTIGUE TODOS OS EMPRÉSTIMOS CONSIGNADOS QUE ENVOLVEM FUNCIONÁRIOS DO SENADO E AS EMPRESAS OPERADORAS.

SARNEY GARANTE QUE A AUTORIZAÇÃO PARA QUE SEU NETO JOSÉ ADRIANO OPERASSE EMPRÉSTIMOS CONSIGNADOS É DE 2005, QUANDO ELE, SARNEY, NÃO OCUPAVA NENHUM CARGO NA MESA DIRETORA.

LOC: POR DETERMINAÇÃO DO PRESIDENTE JOSÉ SARNEY, A POLÍCIA DO SENADO FEDERAL ABRIU INQUÉRITO PARA APURAR A DENÚNCIA DE QUE A LIA RAQUEL DE SOUZA FOI NOMEADA PARA O GABINETE DO SENADOR DEMÓSTENES TORRES SEM QUE NEM MESMO O SENADOR SOUBESSE.

O INQUÉRITO FOI ABERTO A PEDIDO DO PRÓPRIO DEMÓSTENES TORRES, DO DEMOCRATAS DE GOIÁS, POR MEIO DE DENÚNCIA-CRIME. AS PRIMEIRAS TESTEMUNHAS PRESTAM DEPOIMENTO NESTA TERÇA-FEIRA, E NO DIA SEGUINTE SERÁ A VEZ DA SERVIDORA NOMEADA INDEVIDAMENTE E DO PAI DELA, VALDÉQUE SOUZA. A REPORTAGEM É DE ROGÉRIO DY LA FUENTE.

[Comentário 05 = áudio de matéria realizada pelo repórter Rogério](#)

Téc - (0629D01 - T: 0'00" – Rogério/PedroCarvalho/Demóstenes – Inquérito – Operador) (CORTA) O inquérito vai apurar como foi possível Lia Raquel Monturil Vaz de Souza ter sido nomeada por 14 dias para um cargo em comissão no gabinete do senador do democratas de Goiás sem que ele tivesse conhecimento disso. Demóstenes Torres acusa o ex-diretor-geral Agaciel Maia de praticar crime de improbidade administrativa. (0629Demóstenes_Inquérito) *“Com o aparecimento do meu caso, quando ele usurpou uma função sem o meu conhecimento, passou ali a concretização do que todo mundo pensava. Ou seja, ele cometeu um crime contra a administração pública.”* (Rogério) Lia Raquel é filha de outro servidor do Senado, Waldeck Vaz de Souza, que foi assessor direto de Agaciel Maia. Depois ela também foi nomeada, sem conhecimento do titular, para trabalhar no gabinete do senador Delcídio Amaral, do PT de mato grosso do sul. O primeiro secretário do Senado, Heráclito Fortes, do Democratas do Piauí, avalia que a investigação pela Polícia do Senado Federal pode implicar em punições severas. (0629Heráclito_Inquérito) *“Esse é um fato grave! Se comprovado, merece punição exemplar, porque foi apropriação indébita de uma função que pertencia exclusivamente ao senador que não tinha conhecimento do seu uso. Esse, talvez seja o fato mais grave, se comprovado, que nós tivemos até agora dessa quantidade de denúncias recebida.”* (CORTA) (Rogério) (CORTA) O prazo para conclusão do inquérito é de 30 dias, mas pode ser prorrogado se for

concedida autorização judicial.

TÉC: VINHETA

EM INSTANTES: IDELI EXIGE QUE EMPRESAS E GOVERNOS ESTADUAIS AJUDEM UNIÃO A ESTIMULAR O CONSUMO E A ECONOMIA:

TEC: PERGUNTE AO SENADOR:.

LOC: A LÍDER DO GOVERNO NO CONGRESSO, SENADORA IDELI SALVATTI, DO PT DE SANTA DE CATARINA, EXIGIU QUE EMPRESÁRIOS E GOVERNOS ESTADUAIS TAMBÉM ADOTEM MEDIDAS PARA REDUZIR IMPOSTOS E REPASSAR AO CONSUMIDOR AS REDUÇÕES TRIBUTÁRIAS DO GOVERNO FEDERAL.

IDELI DISSE QUE A UNIÃO FEZ A SUA PARTE, AO REDUZIR O IMPOSTO SOBRE PRODUTOS INDUSTRIALIZADOS, O IPI, SOBRE AUTOMÓVEIS, MOTOS, SOBRE A CHAMADA LINHA BRANCA DE ELETRODOMÉSTICOS E SOBRE MATERIAL DE CONSTRUÇÃO.

MAS A REDUÇÃO NÃO CHEGA AO CONSUMIDOR, PORQUE AS EMPRESAS AUMENTAM SUAS MARGENS DE LUCRO.

ALÉM DISSO, OS ESTADOS ADOTARAM O SISTEMA DE SUBSTITUIÇÃO TRIBUTÁRIA, COM O AUMENTO DA COBRANÇA DO ICMS NA ORIGEM DA PRODUÇÃO. POR ISSO, O ICMS ESTÁ BATENDO RECORDES DE ARRECADAÇÃO.

Comentário 06 = áudio extraído de pronunciamento em plenário (SONORA)¹⁵

TÉC: (0629V04 – Ideli – 39”) *“não é possível que nós não tenhamos como chamar o confaz, conselho dos secretários da fazenda de todos os estados, para uma atividade na comissão de assuntos econômicos e perguntar por que os estados estão na contramão, fazendo a substituição tributária sem redução de carga, sem redução dos percentuais do icms. vejam que tipo de resultado nós poderíamos estar tendo se o icms, por exemplo, da geladeira, cujo ipi baixou de 20% para 10%, pudesse ser reduzido de 18% para 9%, para a metade. o resultado*

¹⁵ SONORA: áudio gravado somente com a fala do senador sem a intervenção do repórter

seria fantástico: nós teríamos muito mais vendas, muito mais produção, muito mais emprego”.

LOC: AS VOTAÇÕES NO SENADO PODERÃO SER RETOMADAS NESTA SEMANA. HÁ TRÊS MEDIDAS PROVISÓRIAS QUE TRANCAM A PAUTA.

A PRIMEIRA DELAS É A QUE CONCEDE BENEFÍCIOS FISCAIS PARA AS EMPRESAS QUE PARTICIPAREM DO PROGRAMA HABITACIONAL “MINHA CASA MINHA VIDA”, QUE PREVÊ A CONSTRUÇÃO DE UM MILHÃO DE CASAS POPULARES.

LOC: O SENADOR PAULO PAIM, DO PT DO RIO GRANDE DO SUL, ESTÁ PREOCUPADO COM O AVANÇO DA GRIPE SUÍNA NO ESTADO, ONDE OCORREU A PRIMEIRA MORTE PELO VIRUS, ALÉM DE SETENTA E SEIS CASOS CONFIRMADOS E NOVENTA E SETE SOB SUSPEITA.

Comentário 07 = áudio extraído de pronunciamento em plenário (SONORA)

TÉC: (00629V01 – Paim – 23”) *“mesmo sabendo que o ministério e a secretaria da saúde do meu estado estejam fazendo um trabalho sério e comprometido com a saúde do nosso povo, é importante ficarmos mais do que nunca num estado de alerta, redobrando toda a atenção para o combate à gripe a, que alguns dizem que virou uma pandemia a nível de mundo e é um fato que preocupa a todos nós”.*

TEC: VINHETA

**LOC: JORNAL DO SENADO. APRESENTAÇÃO: _____ E
_____. ÁUDIO: _____.**

LOC: ACOMPANHE, AGORA, AS NOTÍCIAS DA CÂMARA DOS DEPUTADOS. BOA NOITE.

LOC: BOA NOITE.

ANEXO C

Exemplo de um pacote SIP do modelo OAIS utilizando no formato METS

```

manifest.xml
21
22
23 <!-- 1 - METADADOS DE CABEÇALHO -->
24
25 [-] <METS:metsHdr CREATEDATE="2009-07-05T21:30:00" LASTMODDATE="2009-07-05T21:30:00" >
26 [-]   <METS:agent ROLE="CREATOR" TYPE="INDIVIDUAL">
27     <METS:name>Edgar Benicio Rosa</METS:name>
28   </METS:agent>
29 </METS:metsHdr>
30
31 <!-- 2 - METADADOS DESCRITIVOS - PADRÃO DUBLIN CORE (dc)-->
32
33 [-] <METS:dmdSec ID="EX_VB_01">
34 [-]   <METS:mdWrap MDTYPE="DC" LABEL="Dublin Core Metadata">
35 [-]     <METS:xmlData>
36 [-]       <dc:dc>
37         <dc:title>Voz do Brasil</dc:title>
38         <dc:creator>Equipe da Voz do Brasil </dc:creator>
39         <dc:Subject>Voz do Brasil</dc:Subject>
40         <dc:Subject>Jornal do Senado</dc:Subject>
41         <dc:description>Roteiro e audio dos 10 minutos destinados ao Senado Federal no programa "Voz do
Brasil" </dc:description>
42         <dc:description>Arthur Virgilio anuncia que PSDB vai representar contra o Presidente José
Sarney</dc:description>
43         <dc:description>Wellington Salgado diz que Presidente Sarney está sendo vítima de
injustiça</dc:description>
44         <dc:description>Líder do governo exige que empresas e governos estaduais e municipais ajudem a
União a estimular a economia</dc:description>
45         <dc:publisher>Senado Federal </dc:publisher>
46         <dc:contributor> Radio Senado </dc:contributor>
47         <dc:date>2009-06-29</dc:date>
48         <dc:type>text; sound</dc:type>
49         <dc:format>pdf; mp3 </dc:format>
50         <dc:identifier>VB-20090629</dc:identifier>
51         <dc:source>gravacoes de audio de plenário e comissoes, materias e sonoras</dc:source>
52         <dc:language>pt </dc:language>
53         <dc:coverage>br </dc:coverage>
54         <dc:rights>Senado Federal </dc:rights>
55       </dc:dc>
56     </METS:xmlData>
57   </METS:mdWrap>
58 </METS:dmdSec>
59
60 <!-- 3 - METADADOS ADMINISTRATIVOS - TÉCNICOS DE AUDIO - PADRÃO AUDIOMD (amd) -->
61 [-] <METS:amdSec>
62 [-]   <METS:techMD ID="FMT01">
63 [-]     <METS:mdWrap MDTYPE="OTHER" LABEL="FORMATO DO AUDIO">
64 [-]       <METS:xmlData>
65 [-]         <amd:AUDIOMD ANALOGDIGITALFLAG="FileDigital">
66           <amd:audio_data_encoding>mp3</amd:audio_data_encoding>
67           <amd:bits_per_sample>16</amd:bits_per_sample>
68           <amd:data_rate>96</amd:data_rate>
69           <amd:data_rate_mode>Fixed</amd:data_rate_mode>
70           <amd:duration>10:00</amd:duration>
71           <amd:num_channels>1</amd:num_channels>
72           <amd:sampling_frequency>44.1</amd:sampling_frequency>
73           <amd:sound_field>Monoaural</amd:sound_field>
74         </amd:AUDIOMD>
75       </METS:xmlData>
76     </METS:mdWrap>
77   </METS:techMD>
78

```

```

79
80 [-] <METS:rightsMD ID="RMD1">
81 [-]   <METS:mdWrap MDTYPE="OTHER" LABEL="DIREITOS AUTORAIS">
82 [-]     <METS:xmlData>
83
84 [-]       <rights:RightsDeclarationMD RIGHTSCATEGORY="PUBLIC DOMAIN">
85 [-]         <rights:RightsDeclaration>
86           Este objeto está disponível para uso público podendo ser reproduzido sem autorização.
87         </rights:RightsDeclaration>
88
89 [-]         <rights:RightsHolder>
90           <rights:RightsHolderName>Senado Federal</rights:RightsHolderName>
91 [-]           <rights:RightsHolderContact>
92             <rights:RightsHolderContactDesignation>Rádio
Senado</rights:RightsHolderContactDesignation>
93             <rights:RightsHolderContactAddress>Praça dos Três Poderes. Anexo II - Bloco B - Brasília -
DF. CEP:70.165-900</rights:RightsHolderContactAddress>
94             <rights:RightsHolderContactEmail>radio@senado.gov.br</rights:RightsHolderContactEmail>
95             </rights:RightsHolderContact>
96           </rights:RightsHolder>
97
98         </rights:RightsDeclarationMD>
99
100       </METS:xmlData>
101     </METS:mdWrap>
102   </METS:rightsMD>
103
104 </METS:amdSec>
105
106 <!-- 4 - SEÇÃO DE ARQUIVOS - PADRÃO AUDIO (AUDIOMD), TEXTO (TEI), VIDEO (VIDEOMD), IMAGEM (MIX) -->
107
108 [-] <METS:fileSec>
109
110 [-]   <METS:fileGrp ID="ROTEIRO" USE="Transcription">
111 [-]     <METS:file ID="FILE001" MIMETYPE="text/pdf" SIZE="257537" CREATED="2001-06-10T10:00:00">
112       <METS:FLocat LOCTYPE="URL" xlink:href="../O20090629-001/VB-20090629.pdf"></METS:FLocat>
113     </METS:file>
114   </METS:fileGrp>
115
116 [-]   <METS:fileGrp ID="AUDIO" USE="sound recording">
117 [-]     <METS:file ID="FILE002" MIMETYPE="audio/mpeg" SIZE="64232836" CREATED="2001-05-18T10:00:00"
ADMID="FMT01">
118       <METS:FLocat LOCTYPE="URL" xlink:href="../O20090629-001/VB-20090629.mp3"></METS:FLocat>
119     </METS:file>
120   </METS:fileGrp>
121
122 </METS:fileSec>
123
124
125 <!-- 5 - MAPA ESTRUTURAL -->
126
127 [-] <METS:structMap TYPE="logical">
128
129 [-]   <METS:div ID="div1" LABEL="Voz do Brasil">
130
131     <!-- Jornal completo : roteiro + audio completos -->
132
133 [-]     <METS:div ID="div2" LABEL="Voz do Brasil - roteiro" TYPE="text">
134       <METS:fptr FILEID="FILE001"></METS:fptr>
135     </METS:div>
136
137 [-]     <METS:div ID="div3" LABEL="Voz do Brasil - audio completo" TYPE="sound recording">
138       <METS:fptr FILEID="FILE002"></METS:fptr>
139     </METS:div>
140

```

```

141
142
143
144 [-]
145 [-]
146
147 [-]
148
149
150
151 [-]
152 [-]
153
154 [-]
155
156 [-]
157
158 [-]
159
160 [-]
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181

```

```

<!-- Referências a outros documentos METS que descrevem as matérias e sonoras utilizadas neste Jornal -->
<METS:div ID="div4" LABEL="Sonoras" TYPE="sound recording">
  <METS:mptr LOCTYPE="URL" xlink:href="../O20090629-002/O20090629-002.zip"
    xlink:title="Teaser Sen. Arthur Virgilio"></METS:mptr>
  <METS:mptr LOCTYPE="URL" xlink:href="../O20090629-003/O20090629-003.zip"
    xlink:title="Teaser Sen. Wellington Salgado"></METS:mptr>
</METS:div>
<METS:div ID="div5" LABEL="Matérias" TYPE="sound recording">
  <METS:mptr LOCTYPE="URL" xlink:href="../O20090629-004/O20090629-004.zip"
    xlink:title="Sen. Arthur Virgilio e Wellington Salgado"></METS:mptr>
  <METS:mptr LOCTYPE="URL" xlink:href="../O20090629-005/O20090629-005.zip"
    xlink:title="Sen. Heráclito Fortes"></METS:mptr>
  <METS:mptr LOCTYPE="URL" xlink:href="../O20090629-006/O20090629-006.zip"
    xlink:title="Sen. Demóstenes Torres"></METS:mptr>
  <METS:mptr LOCTYPE="URL" xlink:href="../O20090629-007/O20090629-007.zip"
    xlink:title="Sen. Ideli"></METS:mptr>
  <METS:mptr LOCTYPE="URL" xlink:href="../O20090629-008/O20090629-008.zip"
    xlink:title="Sen. Paulo Paim"></METS:mptr>
</METS:div>
</METS:div>
</METS:structMap>
<!-- 6 - LIGAÇÕES ESTRUTURAIS -->
  <!-- <structLink> -->
  <!-- </structLink> -->
<!-- 7 - COMPORTAMENTO -->
  <!-- <behaviorSec> -->
  <!-- </behaviorSec> -->
</METS:mets>

```