

Leonardo Mendes Padilha

**Metadados para preservação digital na
organização arquivística**

Belo Horizonte

2010

Leonardo Mendes Padilha

Metadados para preservação digital na organização arquivística

Dissertação apresentada ao Programa de Pós-graduação em Ciência da Informação da Escola de Ciência da Informação da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Ciência da Informação.

Área de concentração: Produção, Organização e Utilização da Informação.

Linha de pesquisa: Organização e Uso da Informação

Orientador: Prof. Dr. Mauricio Barcellos Almeida.

Belo Horizonte
Escola de Ciência da Informação

2010

P123p Padilha, Leonardo Mendes
Princípios ontológicos no apoio a organização arquivística: uma proposta de estágio terminológico sobre metadados para preservação digital / Leonardo Mendes Padilha. Belo Horizonte: Universidade Federal de Minas Gerais, 2010.

139 f.: il. Enc.

Orientador: Professor Maurício Barcellos de Almeida
Inclui anexos.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

Referências: f. 89 – 91

1. Ciência da Informação – Teses 2. Arquivologia – Teses 3. Documentos Eletrônicos – Teses 4. Ontologia (Representação do Conhecimento) – Teses 5. Preservação Digital – Teses 6. Metadados – Teses 7. Digital Preservation – Teses I. Título II. Almeida, Maurício Barcellos III. Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CDU: 651.5



UFMG

Universidade Federal de Minas Gerais
Escola de Ciência da Informação
Programa de Pós-Graduação em Ciência da Informação

FOLHA DE APROVAÇÃO

"METADADOS PARA PRESERVAÇÃO DIGITAL NA ORGANIZAÇÃO ARQUIVÍSTICA"

Leonardo Mendes Padilha

Dissertação submetida à Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Minas Gerais, como parte dos requisitos à obtenção do título de **"Mestre em Ciência da Informação"**, linha de pesquisa **"Gestão da Informação e do Conhecimento - GIC"**.

Dissertação aprovada em: 15 de dezembro de 2010.

Por:

Prof. Dr. Mauricio Barcellos Almeida (por videoconferência)-ECI/UFMG (Orientador)

Profa. Dra. Cíntia de Azevedo Lourenço - ECI/UFMG

Profa. Dra. Marta Macedo Kerr Pinheiro - ECI/UFMG

Prof. Dr. Marcello Peixoto Bax - ECI/UFMG

Aprovada pelo Colegiado do PPGCI

Profa. Gercina Ângela B. O. Lima
Coordenadora

Versão final Aprovada por

Prof. Mauricio Barcellos Almeida
Orientador



UFMG

Universidade Federal de Minas Gerais
Escola de Ciência da Informação
Programa de Pós-Graduação em Ciência da Informação

ATA DA DEFESA DE DISSERTAÇÃO DE **LEONARDO MENDES PADILHA**,
matrícula: 2008659687

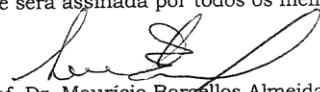
Às 14:00 horas do dia 15 de dezembro de 2010, reuniu-se no auditório 04 do ICEX/UFMG, a Comissão Examinadora aprovada pelo Colegiado do Programa de Pós-Graduação em Ciência da Informação da UFMG em 02/12/2010, para julgar, em exame final, o trabalho intitulado **Metadados para preservação digital na organização arquivística**, requisito final para obtenção do Grau de MESTRE em CIÊNCIA DA INFORMAÇÃO, área de concentração: Produção, Organização e Utilização da Informação, Linha de Pesquisa: Gestão da Informação e do Conhecimento - GIC. Abrindo a sessão, o Presidente da Comissão, Prof. Dr. Maurício Barcellos Almeida (por videoconferência), após dar conhecimento aos presentes do teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Logo após, a Comissão se reuniu sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações:

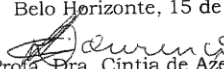
| | |
|--|----------|
| Prof. Dr. Maurício Barcellos Almeida (por videoconferência) - Orientador | APROVADO |
| Profa. Dra. Cíntia de Azevedo Lourenço | APROVADO |
| Profa. Dra. Marta Macedo Kerr Pinheiro | APROVADO |
| Prof. Dr. Marcello Peixoto Bax | APROVADO |

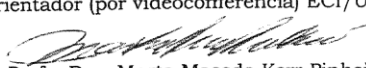
Pelas indicações, o candidato foi considerado APROVADO.

O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a sessão, da qual foi lavrada a presente ATA que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 15 de dezembro de 2010.

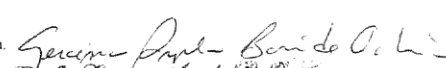

Prof. Dr. Maurício Barcellos Almeida
Orientador (por videoconferência) ECI/UFMG


Profa. Dra. Cíntia de Azevedo Lourenço
ECI/UFMG


Profa. Dra. Marta Macedo Kerr Pinheiro
ECI/UFMG


Prof. Dr. Marcello Peixoto Bax
ECI/UFMG

Obs: Este documento não terá validade sem a assinatura e carimbo da Coordenadora.


Prof. Cíntia Augusta B. D. Lourenço
Coordenadora do Programa de Pós-Graduação
em Ciência da Informação - ECI/UFMG

Agradecimentos

Gostaria de agradecer a todos, mas não posso. Não que não queira, mas porque dentre minhas limitações está a de nunca me lembrar de tudo ou de todos que preciso.

Então, faço apenas das palavras o que elas são. Signos, lembranças incompletas.

Mesmo com o pesar de não fazer dessas linhas algo digno me arrisco a dizer que o trabalho, especialmente o trabalho de pesquisa científica é um trabalho conjunto. Quando alguém chega a um local de destaque o fez porque toda uma equipe, um grupo esteve por traz dessa pessoa, ela apenas representa esse grupo.

Esse é o meu caso. Eu apenas represento uma legião de pessoas que me ajudaram nesse trabalho. Começo minhas lembranças por todos que vieram antes de mim. Aqueles que me permitiram ver um pouco além do que minha visão pequena me permitiu. A todos os amigos pesquisadores ou anônimos que me acompanharam por aqui. Peço perdão a vocês por não citar nomes, mas não quero ser ainda mais injusto ao esquecer de alguém.

Agradeço a toda a equipe da Escola de Ciência da Informação da UFMG. Da secretaria, serviços gerais e demais colaboradores que tornam possível nosso trabalho dentro desse ambiente tão bom, tão rico.

Em especial os meus professores que me acompanharam e me instruíram desde a graduação onde ingressei no ano de 2003 nessa mesma escola.

De maneira ainda mais especial agradeço muito ao meu orientar Prof. Maurício Barcelos de Almeida. Ele pode ser chamado de co-autor desse trabalho, mas isso seria pouco, pois sem ele esse trabalho jamais aconteceria. Mais do que a ajuda eu lhe agradeço pela paciência que aliais devo estender a todos os colaboradores dessa escola. Em minha falta de experiência e simplicidade de enxergar o mundo eu dei muito trabalho, pois nunca conseguia fazer as coisas da maneira mais fácil não por descaso, mas por falta de conhecimento. Perdoem-me.

Agradeço muito a minha família, todos eles em especial a meus pais. Se eles não tivessem me trazido até aqui eu nada seria. Em especial meu pai que muito doente não poderá assistir a esse momento a ele dedico esse singelo trabalho. Um trabalho muito pequeno para o universo da Ciência da Informação, mas gigante se consideradas minhas possibilidades pequenas.

SUMÁRIO

| | |
|---|-----------|
| 1) INTRODUÇÃO | 1 |
| 2) FUNDAMENTOS DA ARQUIVÍSTICA | 6 |
| 2.1) Conceitos Básicos da Arquivística | 6 |
| 2.2) Instrumentos de controle e organização arquivística..... | 8 |
| 2.3) Gerenciamento de documentos de arquivo..... | 16 |
| 2.4) Implicações para pesquisa | 21 |
| 3) PRESERVAÇÃO DIGITAL..... | 23 |
| 3.1) Preservação digital no âmbito da Arquivística | 24 |
| 3.2) Requisitos para a preservação digital | 28 |
| 3.3) Técnicas e estratégias para preservação de documentos digitais. | 31 |
| 3.4) Iniciativas para a preservação digital | 36 |
| 3.4) Implicações para a pesquisa | 40 |
| 4) METODOLOGIA DE PESQUISA..... | 42 |
| 4.1) O objeto da pesquisa | 42 |
| 4.2) Descrição da metodologia de pesquisa..... | 44 |
| 5. COLETA DE DADOS..... | 46 |
| 5.1) Metadata Requirements Project | 46 |
| 5.2) DCMI-Dublin Core Metadata Initiative | 50 |
| 5.4) Cedars-CURL Exemplars in Digital Archives Project..... | 59 |
| 5.5) Preservation Metadata for Digital Collections Initiative | 65 |
| 5.6) METS-Metadata Encoding and Transmission Standard | 68 |
| 5.7) NOBRADE – Norma Brasileira de Descrição Arquivística | 69 |
| 5.8) Implicações para a pesquisa | 73 |
| 6. RESULTADOS..... | 74 |
| 7. CONSIDERAÇÕES FINAIS | 80 |
| REFERÊNCIAS | 89 |

LISTA DE FIGURAS

| | |
|--|-----|
| Figura 1: fragmento do site com definição de termos DC | 48 |
| Figura 2: elementos DCMI definindo um livro em formato digital | 49 |
| Figura 3: exemplos de definição semântica de elementos DCM.. | 49 |
| Figura 4: quadro sinótico de resumo das atividades | 50 |
| Figura 5: Tabela de dados da Metadata Requirements Project | 61 |
| Figura 06: exemplo de descrição do termo resumo do padrão <i>Dublin Core</i> | 63 |
| Figura 07: Exemplo de termos relacionado com o legado no padrão Dublin Core | 63 |
| Figura 08: exemplo de descrição de esquemas de codificação de vocabulário do padrão <i>Dublin Core</i> | 64 |
| Figura 09: exemplo de sintaxe de esquemas de classificação de vocabulário do padrão <i>Dublin Core</i> . .. | 64 |
| Figura 10 - exemplo de classe do padrão <i>Dublin Core</i> | 65 |
| Figura 11: exemplo de vocabulário de Tipos do padrão <i>Dublin Core</i> | 65 |
| Figura 12: exemplo de vocabulário de Tipos do padrão <i>Dublin Core</i> | 65 |
| Figura 13: Representação gráfica do pacote de informação proposto pelo modelo OAIS | 67 |
| Figura 14 – Esquema OAIS de arquivamento e disseminação de dados | 68 |
| Figura 15: Entidades Funcionais OAIS..... | 69 |
| Figura 17 – Exemplo de definição de elemento da | 69 |
| Figura 16: Exemplos de Tipos de PDI. Extraído da Iniciativa OAIS | 70 |
| Fig. 17: Exemplo de elemento extraído da Preservation Metadata for Digital Collections Initiative ... | 77 |
| Figura 18: Exemplo de definição Preservation Metadata for Digital Collections Initiative | 78 |
| Figura 19: Termos extraídos da Preservation Metadata for Digital Collections Initiative | 79 |
| Figura 20: Forma compatível do CRM | 120 |
| Figura 21: inserção termos na estrutura do CIDOC | 122 |

Resumo

A utilização e a criação de documentos no contexto digital é uma realidade. Grande parte dos dados produzidos e armazenados atualmente está nesse suporte. A preservação de documentos em formato digital – em especial no âmbito arquivístico, por seu caráter probatório e muitas vezes histórico – é uma questão que se apresenta na pauta das discussões dessa respeitável área do conhecimento. Unindo princípios teóricos da arquivologia e da preservação digital, o presente trabalho destaca a importância dos metadados para preservação digital, discute o caráter interdisciplinar da abordagem e demarca o espaço da arquivologia. Utilizam-se princípios ontológicos para propor o estágio preliminar de organização da informação, denominado no campo de estudo das ontologias por “estágio terminológico”, no qual se reúnem termos representativos de metadados sob um modelo ontológico de abrangência internacional.

Palavras-Chave: ontologia, preservação digital, metadados.

Abstract

The usage and creation of documents in digital contexts is a reality. The majority of data produced and stored around the world is recorded in a digital support. The preservation of digital documents is a challenge, especially within the archival studies scope, in which documents have probative character and historical value. Indeed, this is a relevant question facing the respectable field of Archival Studies. We here approach principles of archival and digital preservation, present considerations about the importance of metadata in the context of digital preservation, and argue about the interdisciplinary bias of the study bounding the Archival Studies presence. We use ontological principles to propose a preliminary information organization stage, named “terminological stage” in the field of ontologies studies, in which are gathered terms that represent digital metadata. Such set of terms is organized under an international disseminated ontological model.

Key World: ontology, digital preservation, metadata

1) INTRODUÇÃO

Diversos campos de pesquisa ocupam-se de estudar a informação, com destaque para a Ciência da Informação. Verifica-se nos últimos 20 ou 30 anos um aumento expressivo na produção e disseminação da informação registrada, seja em documentos ou em sistemas informatizados. A necessidade de lidar com esse volume de dados têm originado pesquisas interdisciplinares que se vale de ferramental de áreas como a Arquivística e a Ciência da Computação. Nesse novo contexto, busca-se assim atender as demandas sociais das instituições e dos profissionais, relativas à informação.

O termo “arquivo” tem diferentes conotações no âmbito de diferentes disciplinas. Do ponto de vista da arquivística, arquivo é o “conjunto de documentos produzidos e acumulados por uma entidade coletiva, pública ou privada [...] no desempenho de suas atividades, independentemente da natureza do suporte” (CONARQ, 2005 p. 26). Nesse contexto, os arquivos são responsáveis pela comprovação das atividades exercidas pelas entidades públicas e privadas. A memória das instituições é preservada através dos arquivos, os quais são fundamentais para entender o processo histórico de um país ou instituição, seja ela pública ou privada.

A importância dos arquivos e de informações relativas à sua produção exige especial atenção sobre a questão da *preservação*. A preocupação com a preservação de documentos de arquivos não é recente, e existe mesmo em relação a registros pré-históricos. Desde a antiguidade documentos e livros são perdidos por desconsideração de aspectos como o contexto de produção, as traduções e a censura. O esforço feito por arqueólogos e estudiosos para dar sentido aos livros antigos será o mesmo exigido das próximas gerações em recuperar registros correntes, caso não sejam preservados. Segundo Lazinger (2001), a despeito de todo o avanço tecnológico, esse tipo de problema pode se repetir: “[...] livros impressos se desintegraram [...] antigos filmes de cinema foram reciclados [...] a história pode se repetir na evolução da Internet - e sua Web.”¹

A popularização do computador pessoal e seu uso como ferramenta de trabalho disseminaram um novo tipo de documento, o documento eletrônico ou digital. Os documentos de arquivo tradicionais, registrados em mídias físicas como o papel ou o microfilme, têm sido substituídos em larga escala por documentos digitais. Documentos produzidos em formato

¹ “The Early printed books decayed into unrecognizable shreds. Many of the oldest cinematic films were recycled for the silver content. Unfortunately, history may repeat itself in the evolution of the internet – and its World Wide Web”.

tradicional são convertidos para o formato digital, enquanto outros, denominados *nato-digitais*, já nascem em tal formato.

A digitalização tem se difundido dentro do contexto institucional e muitas vezes é vista, de maneira equivocada, como solução para todos os problemas de acesso. Uma dificuldade enfrentada para o uso de documentos digitais como documentos com valor de prova legal é a legislação dos países, que ainda vem se adequando a nova realidade. No Brasil, por exemplo, prontuários médicos podem ser substituídos por equivalentes em meio digital mantendo-se a validade legal. Enquanto os prontuários tradicionais possuem temporalidade de vinte anos, os equivalentes digitais devem ser preservados em caráter permanente.

A partir de uma análise preliminar, a manutenção permanente de documentos digitais pode parecer simples, pois a mídia digital tem características ideais: i) seqüências de *bits* podem ser guardadas por longos períodos tempo, sem perda de informação; ii) grandes volumes de documentos são armazenados em pequeno espaço físico; iii) as buscas são rápidas e a recuperação eficiente, o que simplifica a disseminação de documentos digitais via redes. Entretanto, o armazenamento digital é uma questão complexa, caracterizada por três problemas principais: i) a *interpretação dos dados*: a mídia que registra os dados em formato digital não é diretamente legível pelas pessoas; ii) o *hardware*: falhas nos dispositivos ocasionam a perda de dados e os componentes do sistema se tornam obsoletos mais rápido que as mídias; iii) o *software*: a incompatibilidade entre formatos utilizados ao longo dos anos por diferentes editores proprietários torna impossível o intercâmbio de documentos.

Em poucos anos de digitalização e de produção dos documentos nato-digitais percebe-se a emergência de uma nova e complexa questão, a qual se convencionou chamar *preservação digital*, a qual envolve pelo menos as três questões acima citadas.

A questão da *interpretação dos dados* não corresponde apenas a preservar documentos digitais, mas também os meios de acesso aos documentos e os recursos que possibilitaram sua produção. Um documento digital é uma representação binária que para ser lida por pessoas exige tradução e compilação. Esses processos remetem ao estado da arte tecnológico que se torna rapidamente obsoleto. Existem ainda outros agravantes, como por exemplo, o fato do suporte sofrer degradação física: enquanto o suporte digital torna-se instável em cinco anos, o microfilme ou o papel podem durar centenas de anos.

Em relação ao *hardware*, Hedstrom (1997) destaca que são produzidos dispositivos de armazenamento cada vez maiores em espaço e menores em tamanho físico, mas nada tem sido feito para garantir a manutenção dos dados em longo prazo: “trata-se de uma bomba relógio

[...] enquanto tecnologias de armazenamento abundam, as técnicas de preservação em longo prazo não acompanham essa evolução” (HEDESTROM, 1997, p. 02).

Em relação ao *software*, cabe citar, por exemplo, que planilhas eletrônicas produzidas a cerca de quinze anos atrás em muitos casos já estão inacessíveis. Basta uma nova versão de *software* para que o problema se revele. Segundo Lazinger (2001, p. XV) “[...] as origens da era digital estão provavelmente perdidas”².

Projetos de gestão de informação e de documentos devem prever formas de lidar com a preservação digital, considerando que documentos digitais dependem da tecnologia que os acompanha. Mesmo considerando-se a nova realidade, a organização de arquivos digitais herda conceitos da organização arquivística tradicional. Segundo Malheiro (2001), muitas vezes, e de forma deficiente, o profissional de informática substitui o arquivista em suas funções, especialmente em projetos de gestão de documentos digitais. Ainda assim, a questão central da gestão de documentos, independente da mídia utilizada, ainda está relacionada à recuperação da informação. Ainda segundo Malheiro (2001, p.26),

“Parece existir uma tendência natural a uma “inevitável transdisciplinaridade” na medida em que o objeto de trabalho [...] visa estudar as técnicas de recuperação da informação, quanto [...] o estudo da informação por si mesma parece ser comum a diversas disciplinas [...]. Nesse contexto existe uma grande aproximação entre disciplinas como a biblioteconomia e a arquivística de maneira intrínseca [...], além de uma aproximação [...] de disciplinas como a informática [...].”

A tendência transdisciplinar tem se demonstrado, de fato, interdisciplinar à medida que torna comuns objetivos da Ciência da Informação, da Ciência da Computação e da Arquivística, como por exemplo, a necessidade de lidar com a preservação digital. A importância da questão está na consideração de documentos de arquivo de caráter permanente, fundamentais para compreensão das instituições pelas gerações futuras e mesmo para questões legais. Assim, o problema torna-se mais complexo quando o período de preservação exigido é o longo prazo. Por exemplo, no caso de comprovação de aposentadoria, pode ser preciso manter os documentos por até cem anos.

O presente trabalho se insere nesse contexto interdisciplinar e propõe abordar a questão da preservação de documentos em formato digital a partir dos princípios da arquivística. Em função da complexidade e da extensão do assunto, optou-se por estudar os metadados para preservação digital e sua importância no contexto geral de preservação. Abordam-se as principais iniciativas internacionais de metadados para preservação digital para, utilizando princípios ontológicos para organização da informação, propor uma terminologia alinhada com a

² “[...] the origins of the digital era are probably lost.”

ontologia internacional CIDOC-CRM. Tal estágio preliminar, assim organizado, é o primeiro passo para iniciativas futuras, como base para desenvolvimento de instrumentos de auxílio ao arquivista, por exemplo, tesouros e ontologias.

Ontologias são estruturas baseadas em conceitos e relações, representados por uma linguagem em níveis de formalização variáveis. A ontologia permite entendimento do domínio do conhecimento, descoberta das relações, dos conceitos, a natureza dos termos, generalizações e especializações, bem como a criação de modelos. A ênfase do presente trabalho não é produto final, uma ontologia (a qual será considerada em trabalhos futuros), mas sim os princípios ontológicos para organização da informação. Tais princípios são utilizados para organizar termos representativos de metadados para preservação digital. Assim, cabe salientar que descrever aspectos teóricos das ontologias não faz parte dos objetivos do presente trabalho: apenas são utilizados princípios de organização herdados de ontologia internacionalmente consolidada para organização de metadados.

O **objetivo geral** é organizar dados de iniciativas internacionais sobre metadados para preservação digital em uma estrutura de organização da informação preliminar baseada em princípios ontológicos. Para tanto, colocam-se como **objetivos específicos**: i) entender das relações entre a organização arquivística tradicional e as aplicações no meio digital, como forma de garantir a proveniência e autenticidade de documentos a longo prazo; ii) investigar as principais estratégias utilizadas para preservação de documentos digitais e o uso de metadados nessas estratégias; iii) entender as iniciativas internacionais sobre metadados para preservação digital, de forma a obter termos candidatos para o estágio preliminar terminológico proposto.

Espera-se ao final do trabalho responder as seguintes questões de pesquisa: i) a utilização de metadados é efetiva como forma de garantir a preservação dos documentos digitais a longo prazo? ii) quais as relações entre a arquivística tradicional e a nova realidade que são benéficas para a preservação de documentos digitais a longo prazo? iii) a inexistência de conjuntos de metadados inviabiliza a preservação? iv) o uso de princípios ontológicos para organização de termos representativos de metadados pode auxiliar na construção de instrumentos que auxiliem na preservação de documentos digitais?

O restante do presente trabalho está organizado conforme segue: a Seção Dois apresenta fundamentos da arquivista: conceitos, instrumentos e gestão; a Seção Três apresenta características, requisitos e técnicas para a preservação de documentos digitais a longo prazo; a Seção Quatro apresenta a metodologia proposta; a Seção Cinco apresenta os dados coletados das diversas iniciativas; a Seção Seis apresenta o agrupamento terminológico obtido e sua inserção

no âmbito do estágio preliminar de desenvolvimento de uma ontologia em iniciativas futuras. Finalmente, a Seção Sete apresenta considerações finais e perspectivas para trabalhos futuros. Cabe aqui uma observação para informar que como as iniciativas pesquisadas, assim como a ontologia citada, estão em língua inglesa. Em muitos casos não se traduziu o termo para o português, na crença de que o significado se mantém apesar da mudança da linguagem.

2) FUNDAMENTOS DA ARQUIVÍSTICA

Arquivos são formas de organizar a informação relevante registrada na sociedade, em sua prática corrente e ao longo do tempo. Segundo Schellenberg (1956) são três as razões para a existência dos arquivos: i) a necessidade de eficácia governamental, ii) razões de ordem cultural, iii) razões de interesse pessoal. Na formação de arquivos, o contexto de produção dos registros é fundamental. Tal contexto é garantido pelos princípios de proveniência; que determina que cada fundo de arquivo deve ser organizado de maneira independente, ou seja em separado de documentos oriundos de outros órgãos, secretarias ou empresas. E da organicidade, onde os documentos devem ser ordenados segundo a estrutura funcional que os gerou ou recebeu. Além disso, o uso de instrumentos de descrição dos documentos de acervos em todos os seus níveis é fundamental para que no futuro, seja possível compreender o contexto da geração atual.

Na presente seção, apresenta-se uma abordagem teórica da arquivística a qual está organizada da seguinte forma: na seção 2.1 os conceitos gerais; a seção 2.2 apresenta os instrumentos de controle e organização de arquivos; e na seção 2.3 o gerenciamento de documentos de arquivo. Ao longo do capítulo, e sempre que relevante, é apresentado um paralelo entre as características do documento de arquivo tradicional e aquele em formato digital.

2.1) Conceitos Básicos da Arquivística

Essa seção apresenta uma visão geral da arquivística. Iniciando pela abordagem histórica, desde o surgimento da escrita até os dias de hoje, com ênfase na transição para documentos eletrônicos. Ressalta-se a natureza dos arquivos e os principais conceitos relacionados, desde a criação ou recebimento dos documentos até o descarte ou guarda permanente dos mesmos. Além disso, trata-se das funções administrativas inerentes aos processos de transferência, recolhimento, protocolo e outros fundamentais para o andamento das atividades documentais.

Negreiros (2007) separa duas ênfases para estudo da história da Arquivologia: a primeira está relacionada a história dos arquivos e a segunda trata dos arquivos enquanto instituições. Para os objetivos desse trabalho, a segunda ênfase é de maior interesse. Schellenberg (1956) se refere ao surgimento dos arquivos enquanto instituições na antiga civilização grega. O autor continua com o desenvolvimento dos arquivos na Idade Média em detrimento ao declínio das civilizações antigas. Para os nossos objetivos, considera-se relevante o período do século XIX, na Europa. Na França, a Arquivologia propôs a criação do princípio de

respeito aos fundos, ou seja, de que os documentos deveriam ser organizados segundo os fundos que os geraram. Esse ato é considerado como marco do princípio de proveniência, o qual ainda é a base da arquivística moderna.

Os arquivos passaram a ser organizados segundo os fundos que os geraram, eliminando a anarquia anterior que considerava a organização por assunto. É fundamental o respeito pela organicidade ou arranjo da instituição que produziu os documentos organizando-os segundo um fundo, classe e subclasses que estão subordinadas a hierarquia funcional da instituição que os criou. Deve-se respeitar a ordem natural dos documentos produzidos segundo a organização administrativa da instituição: esse é o princípio da Santidade.

Schellenberg (1956) destaca três elementos fundamentais na caracterização dos arquivos, os quais antecedem as próprias definições de arquivo, pois tratam da própria razão de existir dos mesmos:

- A razão pela qual os materiais foram criados: para ser documento de arquivo o registro deve ter sido criado ou recebido por ato administrativo, ou seja, ele é a prova material de que aquele ato foi realizado;
- Os valores pelos quais os documentos são guardados: documentos são guardados por seu valor primário, ou seja, pelo valor de prova de alguma atividade; ou por seu valor secundário de informação, cultural, histórico, etc.
- A custódia.

Tais razões originam princípios básicos para arquivos. Especialmente quando se lida com documentos digitais, trabalha-se com o conceito de gerenciamento de arquivos, entendidos aqui tanto como a instituição ou setor de guarda, como o próprio documento arquivístico. Esse processo, também chamado de *gestão documental*, envolve mapeamento de processos, criação de organogramas dos setores da instituição e tabelas de temporalidade e destinação de documentos, além de rigoroso controle de tramitação; inclui ainda mecanismos que determinam permissão para visualização, modificação, cópias e outras tarefas relacionadas ao manejo documental e ao documento propriamente dito.

Segundo o Arquivo Nacional (2005), o documento arquivístico deve possuir valor de prova dentro dos contextos legal, fiscal, probatório ou administrativo. Deve ainda, estar inserido em uma comunidade específica. Embora a arquivística possua técnicas parecidas com a biblioteconomia e a museologia, uma vez que todas tratam da informação, as três áreas possuem propósitos diferentes e por isso trabalham o objeto de pesquisas em contextos diferentes. Enquanto a biblioteconomia enfatiza a informação contida no suporte, a Arquivologia aborda o

documento enquanto consequência de uma ação, resultado de um processo. Nesse sentido o primeiro valor de um documento de arquivo é sempre comprobatório e não informativo.

Cabe destacar ainda que os documentos de arquivo possuem características únicas. A Constituição Federal do Brasil considera que qualquer dado pessoal é sigiloso, de forma que a intimidade das pessoas seja preservada. Assim, qualquer documento com dados de um cidadão ganha *status* de sigiloso e deve ser guardado de maneira restrita em relação a grupos que não aqueles que lidam com a documentação por necessidade inerente a função exercida. Segundo o CONARQ (2005) o documento sigiloso é: “documento que pela natureza de seu conteúdo sofre restrição de acesso”. Até mesmo no momento do descarte de documentos o sigilo de ser preservado: os documentos são destruídos antes do descarte. Então, pode se afirmar que documentos de arquivos são diferenciados dos demais (THOMAZ, 2004, p. 79):

“O documento considerado de forma genérica pode ou não atender a definição de documento de arquivo, sendo que, para isso, deverá fazer parte de um conjunto de documentos que guardarão entre si relações específicas, relações essas que transformarão o conjunto de documentos em arquivo. Cada um dos documentos que constituem o arquivo passa, em função dessas relações, a ser denominado documento de arquivo.”

Apresentados os conceitos básicos sobre arquivos, considera-se que a manipulação de documentos passíveis de preservação deve ocorrer sobre a ótica da arquivística, pois a criação dos instrumentos vai determinar quais e como os documentos devem ser armazenados e conservados. Tais instrumentos são descritos na seção 2.2.

2.2) Instrumentos de controle e organização arquivística.

Diversos instrumentos são utilizados para controle, organização e manutenção de arquivos, tanto aqueles compostos por documentos de arquivo em meio físico, quanto digital. A presente seção descreve algumas das principais características e definições relativas aos instrumentos. Dentre os instrumentos cabe citar: documento, arquivo, documento de arquivo, idade do arquivo, transferência, Tabela de Temporalidade, valor primário, valor secundário, fundo arquivístico, descarte, Plano de Classificação, dentre outros. Para conceituar talvez o primeiro e mais importante conceito, Schellenberg (1956, p. 41) define arquivos como

os documentos de qualquer instituição pública ou privada que hajam sido considerados de valor, merecendo preservação permanente para fins de referência e de pesquisa e que hajam sido depositados ou selecionados para depósito, num arquivo de custódia permanente.

O autor ainda define documento (SCHELLENBERG, 2002, p. 41) como,

todos os livros, papéis, mapas, fotografias ou outras espécies documentárias independentemente de sua apresentação física ou características, expedidos ou recebidos

por qualquer entidade pública ou privada no exercício de seus encargos legais ou em função das suas atividades e preservados ou depositados pela aquela entidade ou por seus legítimos sucessores como prova de suas funções, sua política, decisões, métodos, operações ou outras atividades, ou em virtude do valor informativo dos dados nele contidos.

O CONARQ (2005, p. 26) utiliza o mesmo termo para definir documentos de arquivo, instituição arquivo e instalações e móveis de guarda de documentos:

- *Documento de arquivo*: conjunto de documentos produzidos e acumulados por uma entidade coletiva, pública ou privada, pessoa ou família, no desempenho de suas atividades, independente da natureza do suporte;
- *Instituição arquivo*: instituição ou serviço que tem por finalidade a custódia, o processamento técnico, a conservação e o acesso a documentos;
- *Instalações*: locais onde funcionam arquivos;
- *Móveis de guarda de documentos*: móvel e utensílios destinados a arquivos.

Percebe-se que nem todo documento é documento de arquivo. Tal fato é muito importante, pois no âmbito da arquivística, somente são preservados documentos com características arquivísticas. Um conceito fundamental na preservação de documentos, especialmente os digitais, é a aquisição de documentos que possuem valor além daquele diretamente relacionado com sua criação, o valor probatório. Documentos de guarda permanente devem possuir um valor secundário, ou seja, histórico ou informativo.

A partir da segunda metade do século XX, o crescimento do volume de arquivos e a necessidade de espaço, bem como de organização eficiente gerou o conceito das *Três Idades* do documento na arquivologia. Esse conceito também é conhecido como *Ciclo Vital* dos documentos. Nesse contexto, os documentos são separados segundo seu valor dentro do contexto organizacional em três fases: *corrente*, *intermediária* e *permanente*. O CONARQ define ³:

- *Arquivo corrente*: conjunto de documentos, em tramitação ou não, que em função de seu valor primário, é objeto de consultas freqüentes pela entidade que o produziu, a quem compete a sua administração;
- *Arquivo Intermediário*: conjunto de documentos, em tramitação ou não, que em função de seu valor primário, é objeto de consultas menos freqüentes pela entidade que o produziu, a quem compete a sua administração. Em função da legislação ou de outros fatores peculiares da administração, a quem compete sua guarda esses

³ Dicionário Brasileiro de Terminologia Arquivística.

documentos devem ser armazenados em espaço distinto e menos nobre, mas não descartados;

- *Arquivo Permanente*: conjunto de documentos preservados em caráter definitivo em função de seu valor.

A definição das Três Idades é essencial, a medida que a acumulação de documentos cresce no meio físico e mais ainda no meio digital. No segundo caso, os documentos podem ter sua idade alterada, segundo sua necessidade de uso e acesso. Podem receber transferência dos computadores para computadores-servidores, ou mesmo recolhimento para mídias móveis encaminhadas para o setor encarregado da preservação digital.

A *transferência*, ou seja, a passagem do documento da fase corrente para a intermediária; e o seu recolhimento, ou seja, a passagem do documento da fase intermediária para guarda permanente devem ser definidos formalmente através de instrumentos oficiais que minimizem decisões pessoais uniformizando os critérios e políticas de arquivos nas instituições. Além dos critérios legais previstos para guarda ou descarte de documentos, critérios internos de cada instituição são considerados. O valor primário ou secundário de cada documento é que determina a transferência, recolhimento, tempo de guarda e descarte. Embora esse processo possa ser realizado de maneira automatizada, especialmente no caso dos arquivos digitais, a intervenção do profissional ainda é fundamental para a elaboração dos instrumentos.

O *valor primário* de um documento, seu valor de prova, o mantém no arquivo corrente. No caso de necessidade de espaço, pode ser transferido para o arquivo intermediário. Segundo Schellenberg (2002), o valor primário dos documentos é determinado pelos funcionários do setor que o criaram, em conjunto com o arquivista, pois sozinhas nenhuma das partes está preparada para tal juízo. O *valor secundário* dos documentos é determinado, em última instância, pelo arquivista. Esse fato caracteriza a autoridade do arquivista no arquivo permanente, ainda que conte com o auxílio de outros profissionais. Segundo SCHELLENBERG (2002, p. 58):

[...] os arquivistas devem ter a responsabilidade final no julgamento dos valores secundários dos documentos, quer sejam preservados como prova da evolução organizacional ou funcional de um órgão, que sejam preservados por sua informação social, econômica ou qualquer outra.

Outro princípio fundamental da arquivística é o *princípio da proveniência*. Embora questionado por alguns autores esse princípio continua válido e bastante atual. Marca a independência da Arquivologia como disciplina e traz um paradigma próprio, o qual trata a organização da documentação institucional sem depender de outras disciplinas. Segundo o

Dicionário Brasileiro de terminologia Arquivística (ARQUIVO NACIONAL, 2005, p. 135) esse princípio é definido como:

Princípio básico da arquivologia segundo o qual o arquivo produzido por uma entidade coletiva, pessoa ou família não deve ser misturado aos de outras entidades produtoras. Também chamado princípio do respeito aos fundos.

Belloto (2006, p. 130) define o princípio da proveniência como o processo de

[...] deixar agrupados, sem misturar a outros, os arquivos (documentos de qualquer natureza) provenientes de uma administração, de um estabelecimento ou de uma pessoa física ou jurídica determinada.

Heymann (1997, p. 2) diz que o princípio corresponde a

[...] a documentação acumulada por determinada instância, seja ela uma pessoa física, pública ou privada deve ser mantida coesa, respeitando-se a individualidade do conjunto sem misturá-lo a documentos de outras origens, na medida em que tem uma singularidade orgânica que lhe confere sentido, por refletir atividades, dinâmica e critérios da instância responsável pela acumulação.

Similar ao princípio da proveniência é o conceito de *fundo arquivístico*. Na verdade, considera-se que são o mesmo princípio em terminologias diferentes. Para o CONARQ (2005, p. 96), fundo é o “conjunto de documentos de uma mesma proveniência” e pode-se dividi-lo em duas classes distintas: o *fundo aberto*, ou seja, o “fundo ao qual podem ser acrescentados novos documentos em função do fato de a entidade produtora continuar em atividade”; ou o *fundo fechado*, aquele “fundo que não recebe acréscimos de documentos, em função de a entidade produtora não se encontrar mais em atividade”. O fundo é a unidade máxima produtora de documentos em determinado contexto. Segundo Schellenberg (2002) o *Princípio da Proveniência* ou de *Respeito aos Fundos*⁴ desdobra-se em dois:

[...] para Schellenberg o primeiro é o princípio de *respect dès founds*, significando que o arranjo dos documentos não pode se desvincular do lugar de onde vieram, devendo estes ser agrupados de acordo com sua origem nos corpos administrativos de onde provieram. O segundo princípio é o relativo á ordem que tinham os documentos, quando na primeira e na segunda idades. [...] seria o princípio também chamado de “santidade” da ordem original.

Um instrumento fundamental para o gerenciamento da informação arquivística é a *Tabela de Temporalidade e Destinação Documental*, capaz de minimizar o impacto das decisões pessoais dentro do arquivo na medida em que determina de maneira pragmática a transferência e o recolhimento dos documentos para o arquivo intermediário e permanente. Segundo o Conarq (2005, p. 158), a Tabela de Temporalidade Documental é o “instrumento de destinação, aprovado por autoridade competente, que determina prazos e condições de guarda tendo em vista a transferência, recolhimento, descarte ou eliminação de documentos”. Deve ser elaborada na

⁴ “*respect dès founds*”

fase de planejamento da organização arquivística e, para tanto, é fundamental que o arquivista conheça profundamente a estrutura da instituição.

Para a elaboração de um bom instrumento é fundamental a criação de uma comissão de avaliação para destinação de documentos, e conseqüente determinação de prazos de guarda. Tal comissão é composta por arquivistas e representantes dos setores envolvidos. A *Tabela de Temporalidade* é instrumento oficial para determinação do tempo de guarda e destinação de documentos e prevê também a transferência para o arquivo intermediário e recolhimento para o permanente, ou descarte. Nesse contexto, uma opção pode ser a digitalização de documentos muito acessados, ou a micro-filmagem, segundo a necessidade de cada caso. Todas essas questões são tratadas pela comissão de avaliação, provendo credibilidade a Tabela de Temporalidade e inserindo os demais profissionais no processo de organização.

Embora profissionais capacitados e dotados com competência técnica devam ser responsáveis pela elaboração de instrumentos de arquivo a participação de uma equipe inserida no cotidiano da documentação é fundamental. Somente os responsáveis conhecem a fundo as atividades e funções das empresas e do estado. Isso permitirá a separação de documentos por fundos e atividades e determinará as relações entre os setores. É comum na esfera pública que sejam produzidas cópias de documentos oficiais para outros setores com o intuito de facilitar a consulta dos mesmos na fase corrente. Nesse caso, por exemplo, deve proceder-se a guarda dos documentos originais e o descarte imediato das cópias após sua utilização imediata evitando assim o acúmulo de documentação desnecessária nas fases intermediária e permanente.

O princípio da *Unicidade* nos arquivos deve ser observado tanto na forma quanto no conteúdo informacional dos registros. No primeiro caso observa-se a diplomática enquanto balizadora da autenticidade documental que provê o conceito de prova. Deve-se observar que apenas o documento original, respeitados os conceitos legais, pode ser considerado sobre o aspecto probatório. No segundo caso deve-se observar que para cada ato existe um documento que o registra segundo princípios diplomáticos e legais, logo não existem dois documentos que ratifiquem um mesmo ato administrativo embora determinadas atividades possam formar um dossiê com diversos documentos, cada ato é representado apenas por um documento que remete ao conceito de *Unicidade*.

Embora possam ser encontradas cópias documentais com valor informativo que auxiliem as atividades diárias nas diversas esferas do governo, observa-se que tais papéis não são caracterizados enquanto documentos de arquivo independente do formato em que se encontrem. O acúmulo excessivo dessas cópias pode aumentar demasiadamente o tamanho do espaço

necessário para guarda e se transferidos para as idades intermediária e permanente podem acarretar desordem e gastos desnecessários para o administrador do arquivo. É comum, especialmente dentro das repartições públicas, que funcionários desejem recolher ou transferir cópias de documentos no intuito de resguardarem-se contra problemas futuros, contudo isso apenas aumenta o volume de documentos para guarda já que as cópias não possuem valor legal, exceto que sejam autenticadas ou copiadas segundo legislação vigente a ainda sim na ausência do original.

O *descarte* se insere no âmbito do processo de destinação de documentos. Trata-se de processo extremamente relevante, em função da constante multiplicação de documentos. É preciso guardar apenas o acervo necessário e pelo tempo necessário. Uma vez definidos os documentos para eliminação, elabora-se a *lista de descarte* com os documentos a serem destruídos uma única vez. Tal lista é submetida ao responsável pelo procedimento segundo as normas estabelecidas em cada empresa. Após a descrição de todos os documentos para descarte, registram-se a origem e o motivo da eliminação.

O *Plano de Classificação de Arquivos* é um instrumento relacionado a estrutura da empresa e a tipologia documental gerada pela mesma. Segundo CONARQ (2005, p. 131), um Plano de Classificação de Arquivos é um:

esquema de distribuição de documentos em classes, de acordo com métodos de arquivamento específicos, elaborado a partir do estudo das estruturas e funções de uma instituição e da análise do arquivo por ela produzido. Expressão geralmente adotada em arquivos correntes.

Diferentemente de outros tipos de classificação que possuem bases filosóficas, segundo Belloto (2006, p. 136) o arranjo nos arquivos está ligado a “organicidade que prevalece na produção e, conseqüentemente, na organização do arquivo”. Usa-se o termo *arranjo*, no âmbito da arquivologia, para delimitar a base da classificação em arquivos. A organicidade é fundamental para que os documentos sejam conservados da maneira como foram criados, e dessa forma possam testemunhar fatos históricos ou servir como provas.

O Plano de Classificação é elaborado pelo arquivista após o estudo do organograma e mapeamento dos processos em cada setor deve levar em conta o arranjo funcional da empresa, ou seja, aquele “arranjo que tem por eixo as funções desempenhadas pela entidade produtora do arquivo” (CONARQ, 2005 p. 37). Divide-se cada fundo em classes, subclasses e tipologias documentais em um esquema de classificação pré-coordenado. Tal esquema permite localizar a documentação e sua ordenação no arquivo. Observa-se ainda o uso de dossiês, especialmente nas atividades fim onde o processo de geração é fundamental para garantia a manutenção do princípio da proveniência.

Embora o arranjo esteja presente em geral apenas na idade permanente, a documentação é organizada nas três idades do arquivo. A diferença está na quantidade de documentos, que tende a ser maior na fase corrente, menor na intermediária e diminuta na permanente. É impossível guardar todos os documentos e assim é feito um processo de seleção que separa os documentos de guarda permanente dos que são eliminados. Todo esse processo é definido pelo manual de procedimentos oficial da instituição.

Além do processo de classificação e arranjo, é fundamental para a organização e recuperação da informação a elaboração da descrição para os documentos. O processo de descrição é definido como: “conjunto de procedimentos que leva em conta os elementos formais e de conteúdo dos documentos para elaboração de instrumentos de pesquisa” CONARQ (2005, p. 66). Separam-se processos *pré-coordenados*, definidos até o momento e *pós-coordenados*, chamados de descritores. Os arquivos permanentes tem conotação histórica, portanto o público alvo é de pesquisadores e historiadores, pessoas que desconhecem a teoria arquivística. Daí a importância dos descritores para a consulta. Existem diversos instrumentos para recuperação da informação em arquivos, como guias, índices e inventários.

Belloto (2006, p. 173) nos traz a seguinte definição: “o processo da descrição consiste na elaboração de instrumentos de pesquisa que possibilitem a identificação, o rastreamento, a localização e a utilização de dados”. O processo de descrição reúne um conjunto de características que identificam a unidade de descrição segundo o nível desejado. A descrição pode dar-se em vários níveis e sub-níveis identificando desde um item documental até a instituição.

Em 2000 foi apresentada a *ISAD-International Standard Archival Description*, traduzida para o português como NGIDA- *Norma Geral Internacional de Descrição Arquivística*. O ponto fundamental da norma é a estrutura hierárquica com campos que possuem vários níveis de detalhamento. A ISAD fundamentou a construção da NOBRADE- *Norma Brasileira de Descrição Arquivística*, adaptada para a realidade brasileira. Em ambos os casos vale-se do princípio da universalidade que determina que os materiais serão organizados do geral para o específico. A NOBRADE (2005, p.12) determina seis níveis e três sub-níveis de descrição:

[...] acervo da Entidade Custeadora (nível 0), Fundo ou Coleção (nível 1), Seção (nível 2), Série (nível 3), Dossiê ou Processo (nível 4), item documental (nível 5). São admitidos como níveis intermediários o acervo da subunidade custeadora (Nível 0,5), a subseção (nível 2,5) e a subsérie (nível 3,5).

Os níveis de descrição são utilizados segundo a necessidade do acervo da instituição. A variação dar-se-á segundo o nível de especificidade desejado. Em alguns casos é preciso

descrever todos os itens documentais, em outros criar um trabalho exaustivo de descrição é desnecessário; algumas vezes pode ser preciso agrupar documentos em dossiês (nível 4), em outros casos é necessário ir ao item documental (nível 5). Muito embora não se encontre formalmente descrito na literatura observa-se que as atividades fim tendem a ser organizadas em dossiês que agrupam os diversos documentos compondo assim o histórico e o contexto de determinada atividade. Já as atividades meio utilizam-se na maioria dos casos, da organização por tipologia já que suas atividades são comuns a grande maioria das empresas e regidas por leis que determinam suas tipologias e tempos de guarda. Em ambos os casos observa-se, acima de tudo, o respeito a organicidade enquanto princípio que norteia os arquivos.

Tão importante quanto os níveis de descrição são as *áreas de identificação* de documentos. A mesma norma prevê oito áreas de identificação de documentos, as quais contextualizam o documento, atribuem responsabilidade ao trabalho de quem o descreve e auxiliam no processo de recuperação. Segundo o CONARQ (2005, p. 18), as áreas para descrição são:

- Área de identificação: registra informação para identificar a unidade de descrição;
- Área de contextualização: registra informação sobre a proveniência e custódia da unidade de descrição;
- Área de conteúdo e estrutura: registra informação sobre o acesso a unidade de descrição;
- Áreas de fontes relacionadas: registra informação sobre fontes que têm relação com a unidade de descrição;
- Área de notas: registra informação sobre o estado de conservação ou informações sobre a unidade de descrição;
- Área de controle da descrição: registra informação sobre como, quando e para quem a descrição foi elaborada;
- Área de pontos de acesso e descrição de assuntos: registra os termos selecionados para localização e recuperação da unidade de descrição.

Cada área de descrição possui campos obrigatórios e opcionais. A descrição é importante para documentos em meio tradicional ou eletrônico. O conjunto de informações descritas auxilia na preservação de documentos, especialmente os digitais. Um aspecto relevante na organização e na preservação de informações arquivísticas é o contexto da documentação, sua relação intrínseca com a instituição e o propósito de sua guarda. Esses dados são descritos pelos

campos estipulados de áreas de documentação, admitindo-se variações segundo o suporte, gênero ou necessidade específica. Dessa forma, parece também passível de uso no caso de documentos em formato digital. Segundo CONARQ (2005, p. 20),

esta norma deve ser aplicada a descrição de qualquer documento, independente de seu suporte ou gênero. Informações específicas para determinados gêneros de documentos podem e devem, sempre que necessário, ser acrescentadas.

No caso da preservação digital dos documentos de cunho arquivístico é necessária a inserção de campos complementares, conforme descreve Negreiros (2007, p. 62):

No ambiente eletrônico, a descrição ganha ainda maior destaque nas funções de preservação dos documentos e de acesso aos mesmos. Como o acesso ao conteúdo dos documentos eletrônicos está condicionado a sua exibição através da ajuda de um aplicativo de *software*, faz-se uso da descrição do documento eletrônico para acessá-lo previamente, ou seja, desde o início do ciclo vital. A descrição se encarregará de determinar as características nas quais foram concebidas os documentos (aplicativo, fonte, tipo de documento etc.)

Todos os instrumentos e características apresentadas ao longo da presente seção são relevantes para documentos digitais. O fato do documento estar em meio eletrônico muda o suporte, mas não o exime das propriedades de documentos arquivísticos, ao contrário, submete-o as mesmas regras. Em muitos casos, como no princípio de proveniência, o conceito passa a ser de mais importância, pois ajuda na contextualização e na autenticação de documentos digitais. Com informações sobre como o documento foi gerado e em que contexto, torna-se simples determinar a autenticidade no momento da recuperação.

Os instrumentos são orientados por normas arquivísticas sob a supervisão de profissional capacitado, o qual interage com equipes interdisciplinares. No caso de documentos digitais, interage-se em geral com um Cientista da Computação, analista ou programador, o qual deve participar conjuntamente do processo. O conjunto de atividades supervisionadas pelo arquivista é denominado de *gerenciamento* e tratado na Seção 2.3.

2.3) Gerenciamento de documentos de arquivo

Essa seção discute conceitos de gerenciamento de documentos de arquivo em suas três idades, incluindo o mapeamento de processos, instrumentos, técnicas, políticas além das implicações desse processo nas instituições e na organização de documentos, tanto físicos quanto digitais. Ao falar sobre gerenciamento dos documentos arquivísticos usa-se expressão quase sinônima, a gestão documental. Segundo Lopes (2007, p. 02) “o uso da ferramenta gestão documental se faz imprescindível no dia-a-dia das organizações [...]”

O foco desse trabalho se dá em documentos, especialmente os arquivísticos. Entretanto, existe um aspecto anterior a organização, seja ela qual for. Existe a definição de

políticas de informação seja para arquivo, bibliotecas ou centros de informação. Tais definições devem preceder os procedimentos organizacionais pois serão diretrizes para a organização em si. A conscientização dos usuários e produtores, disponibilidade da informação correta, recursos adequados para acesso, especialmente da informação digital são fatores fundamentais e básicos dentro de qualquer sistema.

Mais importante que a organização e a disponibilização de sistemas de informação é a maneira como esses interagem com os usuários. Deve-se levar em conta os arranjos empresariais e sociais em que documentos são criados e manuseados e o pluralismo das situações que envolvem diversas perspectivas e atores. O profissional da informação deve entender bem o seu papel tanto na interação com os profissionais especializados de cada área na construção de classificações pré-coordenadas para arquivos correntes, por exemplo, quanto na avaliação dos valores secundários da documentação nos arquivos permanentes.

Dervin (1994) descreve aspectos importantes relacionados com as políticas de informação. A autora destaca que:

- O acesso à boa informação é fundamental para o exercício da boa cidadania;
- Quando é permitido que a informação transite livremente em um mercado aberto, a “verdadeira” ou “a melhor informação” surge naturalmente como nata em leite fresco;
- Que o valor da “boa informação” é tanto que qualquer pessoa racional vai buscá-la e, portanto, disponibilidade é igual a acessibilidade;
- Que a “boa informação” esteja disponível para todos os cidadãos em uma democracia, que não deva haver nenhuma desigualdade;
- Que infelizmente alguns cidadãos tenham menos recursos, e que, portanto, devemos prover formas de acesso à “boa informação” para tais cidadãos.

Muito embora o enfoque social pretendido por Dervin pode-se afirmar que é a política de informação, é parte fundamental dentro de uma instituição, pois vai determinar a maneira como a mesma lida com esse objeto. Empresas podem ser mais ou menos democráticas com seus funcionários e ter necessidades de controles maiores ou menores sobre a informação que produzem, mas certamente a definição de políticas adequadas será determinante para que se alcancem os objetivos propostos.

Pode-se dizer ainda que existe um dualismo central defendido pela autora que remeta a ordem versus caos. É fato que a condução de uma política em níveis ontológicos, epistemológicos e especialmente ideológicos pode determinar a maneira como uma instituição é

conduzida e seu desenvolvimento. O início de todo esse processo de controle e condução informacional dar-se-á no gerenciamento dos documentos. Muitas vezes a análise de uma política de informação pode levar ao desejo de desinformação para que determinados aspectos da realidade não sejam percebidos e certo nível de alienação seja preservado. Os sistemas de informação, especialmente no meio digital onde adquirem maior poder de disseminação e controle, tornam-se padrões que ratificam convenções ainda que informais. Melhor será se esse padrão estiver sob controle e para tanto se necessita de uma política informacional muito bem definida.

O gerenciamento dos documentos arquivísticos começa na fase corrente, em que são definidas as bases para elaboração dos instrumentos discutidos (Seção 2.2), e criam-se mecanismos para controle da transferência, recolhimento e protocolo de documentos. Ao iniciar a organização do arquivo, o profissional responsável pela coordenação do projeto visita cada setor da instituição munido de questionários que identifiquem a localização do setor dentro da instituição, sua principal atividade de acordo com a missão da empresa, a descrição das atividades e das relações com clientes internos e externos e os documentos gerados pelos processos.

O mapeamento é importante para identificação dos processos essenciais e para análise sistêmica das organizações. Para o desenvolvimento do mapeamento analisa-se⁵:

- Fluxo (volume / unidade de tempo);
- Seqüência de atividades;
- Duração do ciclo;
- Dados e informações;
- Pessoas envolvidas;
- Relações e dependências;
- Insumos necessários;

O mapeamento inclui a análise de todas as atividades realizadas pelo setor com ênfase, no caso da arquivística, nos documentos criados ou recebidos no curso dessas atividades. Para cada ato executado dentro de uma empresa ou repartição um documento deve ser criado como fonte de comprovação. O processo é parte do que legalmente garante a autenticidade e fidedignidade as ações. Após a realização do mapeamento o arquivista pode projetar seqüências de processos, sub-processos, atividades, tarefas e através desses esboços entender, documentar e medir os mecanismos. Tal estudo determina as condições de uso, acesso e transferência de

⁵ Adaptado da apresentação “Mapeamento de processos” da Universidade Federal de Santa Catarina.

documentos, em meio físico ou digital. Em resumo, criam-se procedimentos técnicos e de controle para organização e trâmite de documentos entre os setores e nas diferentes idades do arquivo.

Além dos desafios na criação do sistema, tem-se outros em sua manutenção, já que em muitos casos os profissionais dos setores envolvidos não entendem a importância dos processos. Segundo LOPES (2008, p. 4):

As instituições são levadas a investir em seus recursos informacionais, absorvidas pelo fenômeno da globalização, em cujo mercado se sobrepõem seus negócios; não basta apenas produzir a informação e sim possibilitar o intenso uso dela visando o rápido acesso e recuperação da informação; para tanto é necessário fazer uso da gestão documental.

O gerenciamento de arquivos é uma tarefa de suma importância para entendimento das instituições em seu contexto de atuação. Segundo Arantes (1998), através das informações contidas em documentos de arquivo pode-se conhecer o que foi feito no passado, o que está sendo feito no presente e o que será feito no futuro. A legislação nacional dispõe sobre a atividade de gestão documental desde 1991 (BRASIL, 1991, p. 1) como os “[...] conjuntos de procedimentos e operações técnicas referentes a produção, tramitação, uso, avaliação e arquivamento em fase corrente e intermediária, visando sua eliminação ou recolhimento para guarda permanente”.

Do ponto de vista prático, para melhor avaliação, cria-se um roteiro para entrevistas com os facilitadores de cada setor os quais, preferencialmente, devem compor também a comissão para avaliação de documentos. Nesse questionário deve-se considerar, no mínimo, as seguintes informações:

- Nome da instituição, nome do responsável pela coleta dos dados, objetivo do questionário, setor entrevistado, responsável pelas respostas, gerente do setor.
- Questionamentos sobre:
 - Hierarquia do setor no organograma da empresa;
 - Identificação das atividades do setor;
 - Identificação dos processos inerentes às atividades;
 - Conjunto de documentos gerados pelas atividades;
 - Protocolos de transferência de documentos;
 - Tempo das atividades, incluindo tempo de envio e recebimento de resposta a solicitações;
 - Identificação do tempo de permanência dos documentos no setor;

- Identificação dos critérios de transferência ou da necessidade de criação de documentos para arquivo intermediário;
- Identificação dos pontos que precisam ser melhorados e de não conformidade dentro do contexto arquivístico do setor.

Após visita aos setores, cria-se um diagnóstico da situação da instituição do ponto de vista arquivístico e desenvolvem-se os instrumentos necessários para gestão documental, desde a produção do documento até o recolhimento ao arquivo permanente ou descarte. Nesse processo alguns instrumentos fundamentais são: formulários de transferência de documentos, físicos e eletrônicos, entre setores e dos setores para o arquivo intermediário; cronogramas de transferência de documentos; protocolos de envio e recebimento de documentos; tabela de destinação de documentos, que pode ser conjunta com a Tabela de Temporalidade. Esse último instrumento determina transferência, recolhimento ou eliminação dos documentos, assim como a migração deste para outros suportes e necessidades especiais de preservação quando em meio digital ou outros que necessitem de tratamento especial.

LOPES (2007, p. 09) apresenta um estudo em que descreve as seguintes soluções, após o diagnóstico da documentação corrente e permanente de um sistema: i) elaboração de instrumentos de tratamento arquivístico; ii) implementação de metodologia e forma padronizada para documentos; iii) elaboração de manuais de procedimento e de usuários; iv) unificação de sistemas de recuperação de informações; v) aprimoramento e extensão de modelos de tratamento do acervo fotográfico; vi) unificação dos contratos de guarda de documentos; vii) tratamento e transferência dos documentos acumulados nas unidades correntes; viii) conscientização e descentralização de responsabilidades; ix) implementação da gestão documentos em arquivos correntes. Tais soluções não podem ser consideradas modelo para gestão da informação, mas apresentam um panorama razoável dentre as instituições.

Para elaboração de um plano de preservação tradicional, ou digital, todo o processo da organização arquivística deve estar completo. Para preservar é preciso organizar e definir, segundo a política de arquivo da instituição, o que será preservado e por quanto tempo. Por exemplo, documentos do setor de pessoal são guardados por trinta anos ou até mesmo em caráter permanente; documentos contábeis podem ser descartados num prazo de cinco anos. Além disso, deve-se levar em conta o valor secundário dos documentos que é subjetivo em relação as necessidades de cada organização.

Após o diagnóstico é elaborado um plano de arquivo com previsões de custos, pessoal envolvido, cronograma e ações necessárias para a implementação da política de arquivo

para organização e preservação dos registros físicos e digitais. No segundo caso, é necessária uma equipe especializada e interdisciplinar, envolvendo profissionais da informação e tecnologia.

Após a organização arquivística, o responsável pelo centro de documentação está apto a adquirir, com segurança, os documentos selecionados para preservação. A partir do estabelecimento da segurança em um sistema de organização, avança-se para a fase de preservação de registros digitais, conforme será apresentado na Seção Três.

2.4) Implicações para pesquisa

A presente seção apresentou uma breve abordagem teórica a arquivística, descrevendo os conceitos gerais subjacentes a disciplina, os instrumentos de controle e de organização de arquivos, além dos processos de gerenciamento de documentos de arquivo. Sempre que possível, referências e paralelos com arquivos em formato digital foram apresentadas.

Observa-se que os princípios arquivísticos para documentos em papel também são válidos, pelo menos em grande parte para o meio digital. Muitos teóricos relacionam a mudança de suporte como uma revolução, mas esse trabalho não encontrou indício que justificassem tal afirmação. Deve-se observar que o suporte eletrônico traz uma realidade nova que demanda novos parâmetros. Esses devem ser somados aos anteriores na medida necessária a nova realidade.

É muito provável que a arquivologia deva rediscutir alguns conceitos, antes ignorados como análise de assunto ou uma indexação mais profunda, mas isso, provavelmente, não afetará suas bases e marcos. Uma discussão desse nível não se dá pela mudança do suporte, por maior que seja, mas pela análise histórica e filosófica de toda a área. O que se discute são novas necessidades e novas oportunidades em ambos os casos criadas pelo meio digital. Dentro dessa idéia está a preservação digital, uma nova necessidade abordada pelo presente trabalho.

O suporte eletrônico traz possibilidades nunca imaginadas de armazenamento e recuperação de informações e suscita, nesse sentido, novas discussões que vão além do técnico. Até que ponto a ordem original dos documentos será fisicamente mantida no meio digital? É uma das perguntas que podem suscitar novas pesquisas e apontamentos. Os princípios arquivísticos estão intrinsecamente ligados à história, à administração e especialmente as idiosincrasias históricas que, segundo Foucault (2008), parecem mover-se por emergências.

Assim constituiu-se o princípio dos fundos que repete a ordem vigente ainda que essa represente a reprodução dos interesses de determinado grupo.

Como se pode ver as discussões profundas sobre os princípios arquivísticos não se dão pela mudança de suporte, mas pelos movimentos históricos e pela reflexão profunda e acurada das disputas de poder e da formação de cada estado, por isso existem diversas escolas que divergem sob os pontos de vista.

Naturalmente, a mudança do suporte trouxe a necessidade de novas propostas. A natureza do suporte digital em si mesma remete a necessidade de metadados. A discussão mais aprofundada será feita durante o trabalho, contudo deve-se dizer que cada arquivo, considerando o mesmo como a menor unidade dentro do sistema arquivístico, está vinculada a uma complexa rede de software e hardware que não pode ser codificada senão pela presença de metadados.

No meio digital pode-se capturar o documento e exercer um rígido controle sobre o mesmo no momento da sua criação. O que antes significava um esforço hercúleo já que, considerando os níveis de descrição propostos pelo CONARQ tratar um documento no nível 05, ou seja, do item documental significava um grande esforço devido ao elevado número de documentos presentes em determinado dossiê no meio digital isso pode ser feito de maneira automatizada através de sistemas previamente parametrizados pela necessidade de cada organização. Um grande salto na produtividade e no controle da cada atividade.

Como se pode perceber diversos fatores permeiam a discussão arquivística e sua migração para o meio digital. Para o cerne dessa pesquisa importa a discussão sobre os metadados. Especialmente aqueles relacionados com a preservação digital. Optou-se, portanto, em considerar a base da arquivologia tradicional e seus instrumentos descritos no presente capítulo incluindo novas perspectivas segundo a ótica do acesso a longo prazo de documentos digitais.

A importância deste capítulo para a pesquisa reside em duas questões principais. A primeira diz respeito a realizar um estudo interdisciplinar, que envolve assuntos abordados na Arquivística, Ciência da Informação e Ciência da Computação. Cabe assim definir um mínimo da teoria arquivística, a qual se acredita ser fundamental para os objetivos do trabalho. A segunda questão está relacionada justamente a importância da arquivística no âmbito da pesquisa: ao lidar com documentos em formato digital o que muda é apenas o suporte. Dessa forma, é importante entender e aplicar a experiência e as soluções propostas pela disciplina arquivística e aproveitá-la para o contexto dos arquivos em formato digital.

3) PRESERVAÇÃO DIGITAL

Conforme já mencionado, verifica-se um rápido crescimento no volume de informação registrada em forma digital. Os documentos são criados como resultado da digitalização de coleções não digitais, ou como documentos nati-digitais e como parte das atividades rotineiras de uma organização. A preservação digital tem muitos paralelos com a preservação tradicional em matéria de princípios gerais, mas difere no nível operacional. Ainda assim, existe um grande número de colaboradores de uma instituição que desempenha atividades cruciais nos vários estágios do ciclo de vida de um recurso digital, o que torna a tarefa complexa.

É importante abordar o tema política de informação, em especial dentro do ambiente digital. Nesse caso a questão do controle de acesso a documentação que remete a segurança e autenticidade passa de importante à fundamental dentro de qualquer processo. Os temas tratados na seção 2.3 são válidos também dentro do contexto aqui apresentado. O documento deverá ser submetido à auditoria remetendo a estudos da diplomática e tipologia que resultem nos atributos necessários para seu manuseio e recuperação seguros. Segundo Belloto (2006), os estudos de diplomática e tipologia levam a entender o documento desde o seu nascedouro, a compreender o porquê e o como ele é estruturado no momento de sua produção.

Para os objetivos do presente trabalho, é necessário entender as bases para a criação de modelos que, partindo da organização arquivística aliada a um trabalho interdisciplinar com a informática, possam estabelecer o conjunto de requisitos mínimos para preservar registros digitais no longo prazo. Espera-se contribuir para tal desafio com a definição do estágio terminológico preliminar de metadados sobre preservação digital.

Para entender o domínio do conhecimento da preservação digital, esse capítulo discorre sobre preservação digital à longo prazo e está organizado da seguinte forma: a Seção 3.1 apresenta conceitos básicos; a Seção 3.2 apresenta requisitos para preservação digital; as seções 3.3 e 3.4 apresentam técnicas e estratégias para preservação de registros digitais, bem como o CIDOC-CRM, uma ontologia relevante para os objetivos desse trabalho. Dessa forma, buscou-se enfatizar iniciativas e padrões que se alinhem aos objetivos da pesquisa: o entendimento do domínio de conhecimento e a delimitação das iniciativas de metadados para preservação digital, as quais são a base para desenvolvimento da proposta. Finalmente, a Seção 3.4 apresenta as implicações para a pesquisa no capítulo.

3.1) *Preservação digital no âmbito da Arquivística*

Conforme mencionado anteriormente, no contexto da preservação de registros digitais a longo prazo enfatizam-se aqui os documentos arquivísticos, muito embora os conceitos possam ser estendidos para outros tipos de documentos. Segundo Thomaz (2004, p.89), “a primeira preocupação que surge com relação ao documento eletrônico de arquivo, é distinguir sua natureza específica”. É notório que a evolução da tecnologia tem trazido grande parte da documentação para o meio eletrônico, o que leva a necessidade de novas abordagens. Ainda assim, os princípios arquivísticos são fundamentais em processos de acondicionamento e preservação.

A memória das instituições e mesmo das pessoas encontra-se ameaçada em um futuro próximo, uma vez que grande parte está registrada em arquivos digitais, efêmeros e voláteis. Contribui para isso também a falta de preocupação em relação ao problema por parte da maioria das pessoas e instituições. É importante salientar o valor histórico dos registros sistematicamente eliminados, num processo que muitas vezes passa despercebido pela sociedade. Considerando que a maioria das informações produzidas no mundo está em meio digital, a perda sistemática desses registros pode significar um retrocesso social semelhante ao produzido em sociedades sem registros escritos. Segundo Le Goff, (1984, p.16)

Nas sociedades sem escrita, a memória parecia ordenar-se em torno de três grandes interesses: a identidade coletiva do grupo que se funda em certos mitos, mais precisamente nos mitos de origem religiosa; o prestígio das famílias dominantes que se exprimem pelas genealogias; e o saber técnico transmitido por fórmulas práticas fortemente ligadas à magia religiosa.

Destaca-se no processo de registro da memória escrita a invenção da imprensa que aumentou significativamente a quantidade de registros escritos e promoveu uma verdadeira revolução social. No caso dos arquivos, o século XVIII foi a época em que os registros ganharam força e independência, como tratado anteriormente. Assim como a imprensa, a internet e a facilidade de criação de registros digitais estão modificando a maneira como as pessoas se comunicam e produzem registros. A internet trouxe de volta o hábito de escrever que havia sido substituído pelo registro oral com o advento do telefone.

A questão da preservação digital vem sendo debatida há anos por especialistas de diversas áreas. Grupos interdisciplinares com pesquisadores internacionais buscam soluções para o problema que possui gravidade maior que aquela que se apresenta a princípio. Caso o processo não seja rapidamente revertido pode-se chegar a uma etapa da história vaga, perdida para sempre. Segundo um estudo da *University of California / School of Information Management and System* (2003), 92% das informações produzidas no mundo em 2002 foram registradas em

meio magnético. Esse dado é alarmante ao se considerar que a quase totalidade desses dados não recebe qualquer tipo de tratamento visando a preservação à longo prazo e parte significativa diz respeito a documentos científicos ou de valor histórico para a humanidade.

Documentos digitais podem ser modificados com muita facilidade. Além de garantir o princípio da proveniência, é fundamental o controle de todas as alterações e de que toda a tramitação seja rastreada para garantir autenticidade. Outra questão relevante é a dos direitos autorais num ambiente como a Internet. A legislação da maioria dos países não está preparada para lidar com essa realidade e tratados internacionais serão necessários para garantir o direito legal de copiar para preservar.

É comum nos dias de hoje que documentos contêm outros documentos anexados, *links*, conteúdo multimídia, dentre outros. Com isso, são necessários *softwares* ou sistemas operacionais, além de diferentes *hardwares* para ler e interpretar um pequeno conjunto de documentos anexos a um *e-mail*. Esse conjunto constitui uma lógica particular e complexa que exige, por parte do usuário, um conhecimento avançado dos ambientes e da lógica interna de cada contexto. Segundo Thomaz (2004, p. 90) “o documento deixa, portanto, de ser essencialmente o objeto físico para se transformar no objeto lógico”. Os documentos digitais possuem também, além de seu conteúdo, uma série de outros dados, ou metadados, inseridos em seu contexto.

Em teoria, o documento digital não existe até que seja decodificado por padrões pré-estabelecidos, para que possa ser apresentado em um formato inteligível para a percepção humana. Com isso obtêm-se textos, figuras, vídeos. Entretanto, o que parece tornar essa informação relativamente simples, a torna potencialmente complexa: o conjunto binário, denominado *bit stream*, recebe como acréscimos uma série de metadados relativos a sua formatação, *hardware*, sistema operacional, dentre outros. Na verdade cada etapa de processamento de um documento agrega-lhe novos dados, e ao final, o objeto torna-se complexo e a acessibilidade dependente de fatores externos de *software* e *hardware*.

No caso da Internet, além das preocupações citadas, considere-se ainda a alteração de endereços, os chamados *links* quebrados, além do dinamismo característico da Internet, em que o usuário é também editor e interage com o conteúdo modificando-o a todo o momento. Ainda que a maioria dessas informações possua características apenas de referência, o contexto histórico de transição representado por esse universo precisa ser preservado enquanto importante capítulo da história da humanidade. Para Le Goff (1984, p. 41) “a memória constitui, ao mesmo tempo, o material e o objeto”.

Um dos atores importantes para a preservação são os produtores de *software*. No caso dos *softwares* proprietários, não se pode garantir que a empresa divulgará novas versões, ou manterá versões compatíveis, para que mais pessoas adquiram a atualização. Isso remete a busca por soluções de código aberto que, em teoria, possuem chances de criar padrões mundiais, pois proporcionam trabalho conjunto. Além disso, o chamado *software open-source* está livre das idiossincrasias comerciais que trazem em sua raiz a inovação como algo superficial, meramente comercial em suma apenas mais uma novidade mercadológica em detrimento a evolução natural do estado da arte tecnológico.

Os documentos digitais são representações dos documentos reais. Por isso diversos estudos indicam que a preservação de documentos deve enfatizar o acesso. Os modernos documentos possuem características que vão além da informação em si. Planilhas possuem fórmulas, editores de textos possuem recursos de formatação, etc. Rapidamente, um documento torna-se complexo e exige mecanismo de conversão para a linguagem humana que consiga reproduzi-lo. Essa ferramenta vai além de um simples *software* de leitura de um conjunto de caracteres, uma vez que inclui, no mínimo, os metadados que o configuram. Além do *software* original, o sistema operacional e o *hardware*, também o tipo de mídia e a especificidade de cada tipo de documento determinam mudanças nos metadados.

Os metadados são parte integrante e fundamental dos arquivos digitais, para os quais se adota, em geral uma definição simplificada: metadados são dados sobre dados. Embora essa definição possa ser um ponto de partida, surgem novas perspectivas e uma abordagem mais completa para a questão. Metadados para documentos tradicionais correspondem a descrição de determinado registro, a qual conecta o usuário, que busca um documento ou informação e o próprio registro. Nesse sentido, diversos campos são definidos, segundo o contexto da informação, visando realizar buscas com revocação adequada e alta precisão. No caso dos documentos digitais, os metadados ganham novos atributos: além dos dados necessários para a busca dos documentos, os metadados incluem os atributos que informam ao *software* ou sistema operacional os requisitos fundamentais para as operações sobre o registro.

A preservação digital no âmbito da arquivística remeta a controles rígidos. Isso faz referência a um *software* de gestão eletrônica de documentos, ampliada para arquivos. O *Instituto Nacional dos Arquivos de Portugal*, em referência ao documento de modelos de requisitos para a gestão de arquivos eletrônicos (conhecido como MoReq), destaca a diferença básica entre o *software* gerenciador de documentos arquivísticos e o sistema gerenciador de documentos arquivísticos. Enquanto o primeiro trata da gestão de quaisquer documentos, o

segundo trabalha com documentos específicos inseridos no contexto arquivístico, os quais possuem uma série de requisitos que devem ser observados.

Thomaz (2004, p. 56) chama a atenção para a evolução da preservação digital no mundo e especialmente no Brasil, aonde iniciativas vêm sendo tomadas para garantir o acesso em longo prazo ao material. Ao destacar a criação do programa *Sociedade da Informação no Brasil*, a autora ressalta que, embora o documento trate sobre questões legais e insira o registro eletrônico dentro da prática arquivística do Arquivo Nacional, não aborda nenhum aspecto da preservação digital. Lembra ainda que o CONARQ possui a Câmara Técnica de Documentos Eletrônicos que trata, desde 2002, ano da sua criação, da gestão e preservação de documentos eletrônicos. No ano de 2004 a Câmara cria a *Carta de preservação do patrimônio arquivístico digital* e, através de sua resolução de número 20, determina a inserção de documentos digitais em programas de gestão arquivística.

Grandes desafios colocam-se para os arquivistas que lidam com a preservação digital. Os grupos de estudos, com destaque para o *Projeto Interpares*, apontam soluções, mas são unâmines em reconhecer que não existe, no atual estado da arte, tecnologia capaz de proporcionar solução definitiva. Hedstrom (1999, p. 05) aponta quatro áreas prioritárias e fundamentais para pesquisa na área: *mídias de armazenamento, migração, conversão e ferramentas de gerenciamento*. Cabe ainda incluir o *encapsulamento* e os *metadados*. A autora ressalta ainda que os quatro domínios são dependentes e estão ligados a estrutura de preservação. Encontrar soluções nos quatro campos e na integração entre os mesmos é um requisito para proporcionar acesso as gerações futuras, aos documentos digitais produzidos hoje.

A preservação digital está diretamente ligada a gestão de documentos digitais de caráter arquivístico. Os temas são próximos na medida em que documentos de arquivo devem ser preservados e tal preservação exige organização documental, que em sua essência é objeto de trabalho e de estudo da arquivística. Muitos instrumentos da práxis arquivística podem ser usados e até mesmo aperfeiçoados para produção de métodos de preservação de documentos digitais. Nota-se ainda uma semelhança entre a Arquivística e a organização dos arquivos digitais no âmbito da informática. A complexa e elaborada rede de requisitos e processos lógicos para visualização de documentos em meio digital parece estabelecer determinados padrões para documentos de um dado sistema. Além disso, se inseridos num mesmo contexto, formam a repetição de dados que leva a uma tipologia na extensão do arquivo, muito útil para construção de arranjos.

A análise dos conceitos básicos da preservação digital e suas implicações levam a necessidade de definir requisitos para a preservação de registros digitais em longo prazo que serão abordados na Seção 3.2.

3.2) Requisitos para a preservação digital

Para ser efetiva, a atividade de preservação digital exige um conjunto de tarefas, processos e planejamentos, do ponto de vista técnico, organizacional e tecnológico. A presente seção descreve os principais requisitos para a preservação digital e apresenta considerações sobre a aplicação desses requisitos em instituições, em especial em instituições arquivísticas que manipulam informações em meio digital.

Dois fatores principais são importantes para a preservação de documentos tradicionais e digitais. O primeiro fator, externo, de acondicionamento do suporte do documento em condições ideais, segundo necessidades específicas, que considera o local e suas peculiaridades: controle de temperatura, umidade, pragas, ventilação, luz e acesso, dentre outros. E o segundo fator, relativo as características intrínsecas de cada documento. No caso dos documentos digitais, os suportes são os computadores e mídias removíveis. Nesse caso, a influência vem de fatores como as condições do ambiente e das características lógicas dos documentos em si, os quais vão exigir revisões periódicas, cópias de segurança e atualização.

Outro fator importante é o tempo que o documento deverá ser armazenado. Essa avaliação evita equívocos e perda de dados. Determina o que deve ser preservado à longo prazo, bem como os recursos necessários. Em função da tecnologia, um prazo entre três e cinco anos deve ser estipulado para a revisão dos requisitos de preservação. Tanto Fatores físicos quanto lógicos devem ser observados. Isso envolve o planejamento da TI da instituição; possíveis atualizações de software devem ser lembradas para efeito de preservação.

Arquivistas devem participar efetivamente das decisões. A tecnologia é um dos pilares da administração moderna. Tendo como base a informação não pode manter-se alheia as demais disciplinas que lidam com o tratamento da informação. O arquivista, por sua vez, deve ter por objetivo cuidar da documentação digital na mesma proporção em que cuida da documentação física. Voltando-se, em especial, para a acessibilidade e tempo de guarda. Em resumo, a arquivística somada às ações práticas de preservação digital irá garantir o acesso a documentos em longo prazo. A definição de uma *política de preservação digital* com a criação de um setor específico e preferencialmente interdisciplinar é essencial para lidar com as novas questões. O trabalho é parte da política arquivística da instituição.

Segundo Thomaz (2004), inúmeros problemas cercam a preservação digital:

- Carência de políticas de avaliação na produção do documento, o que gera documentos inúteis, uma vez que não se tem noção de sua importância naquele contexto;
- Carência de políticas de descrição;
- Vulnerabilidade física do *hardware* que se danifica e com facilidade;
- Vulnerabilidade lógica do ambiente informático, que compromete a integridade, a autenticidade e a história dos objetos digitais;
- Alta obsolescência tecnológica, que reduz o tempo de reação a mudanças de décadas para anos;
- Alta dependência tecnológica;
- Dificuldade de obtenção de *expertise*, pois a tecnologia para acesso aos objetos digitais requer mão de obra altamente especializada.

Todos os processos descritos acima se resumem na falta de uma política adequada para processos informacionais. É bastante comum, especialmente na esfera pública brasileira, o acúmulo de documentação duplicada até mesmo no meio físico. Esse problema aumenta no ambiente digital onde a facilidade de criação e cópia dos documentos aumenta. A falta de políticas adequadas de descrição também ajuda muito na perda do valor secundário da informação na medida em que descaracteriza o contexto que teve na época de sua criação e no caso da informação digital impede sua recuperação, pois não fornece informações fundamentais para o usuário pesquisador. Some-se a rápida obsolescência da tecnologia e a dificuldade e custo da mão de obra altamente especializada e tem-se um panorama complexo para administração.

Thomaz (2004, p. 118) lembra ainda que é mais difícil manter a integridade dos objetos digitais, em detrimento a seus equivalentes tradicionais, pois

[...] quando um relatório impresso é preservado em seu formato original, todos os seus aspectos são mantidos na presença física [...] Objetos digitais, ao contrário, podem facilmente ser decompostos em seus elementos individuais, sendo necessário maior esforço para preservá-los como um todo.

A preservação digital altera a ênfase da preservação do documento para o acesso do mesmo à longo prazo. Para que um documento seja acessível por muitos anos é fundamental que sejam definidas as condições, objetivos, restrições para todas as partes envolvidas no processo. Para esse trabalho adota-se um conjunto de regras e requisitos.

Um importante requisito é a garantia da integridade lógica. É fundamental que o documento tenha sua autenticidade garantida e, para tanto, deve ser monitorado desde sua

criação até a destinação final. Bullock (1999), como base no OAIS-*Open Archival Information System*⁶, define ações as quais denomina requisitos:

- Delimitar o objeto: identificar, dentre os conjuntos de objetos, aqueles que de fato deve ser preservado;
- Preservar a presença física: em arquivos digitais as séries de 0s e 1s são a base para o significado do objeto digital;
- Preservar o conteúdo: a informação do documento em formato ASCII, a qual é independente da variação de leiaute;
- Preservar a apresentação: incluir características como margens, espaços, cabeçalhos, dentre outras;
- Preservar a funcionalidade: conservar as características multimídia, por exemplo, de hipertexto ou navegação;
- Preservar a autenticidade: garantir que o documento acessado é o mesmo produzido, e que mudanças para preservação não afetam a forma original;
- Acompanhar acesso, transferência ou modificação: garantir rastreabilidade desde a sua criação e garantir a localização da versão correta;
- Preservar o contexto: a dependência do *hardware* e de *software*, além de sua relação com outros objetos e contextos.

Além dos passos descritos, considera-se também o princípio de proveniência e o arranjo original dos arquivos, uma vez que os documentos são de caráter arquivístico. A proveniência não se dará fisicamente, como no caso dos documentos tradicionais, mas de maneira lógica nos metadados que descrevem esse princípio.

A documentação de grandes instituições remete a grande volume de dados. Nesse caso é recomendado usar padrões de arquivos para que as estratégias de preservação, em especial a migração, possam ser realizadas em grande escala. Por exemplo, se todos os arquivos de fotos, usarem a mesma extensão um critério único para migração e atualização do formato pode ser adotado. No caso de formatos múltiplos será necessário desenvolver critérios múltiplos e em muitos casos em conversão de formatos que podem resultar em perda de dados. Bullock (1999) sugere etapas de abordagem à questão:

- Adoção de padrões;
- Desenvolvimento de orientações para a preservação digital;

⁶ Disponível na Internet em <http://public.ccsds.org/>

- Descrição detalhada de cada registro a ser preservado;
- Criação de identificadores únicos para cada documento;
- Criação de *links* dos metadados para o contexto.

A adoção de requisitos básicos concernentes a preservação digital é o primeiro passo de um processo que demanda tempo, estudo e trabalho interdisciplinar. A partir da definição de uma política de gestão documental, seguida da definição de requisitos que considerem as mudanças de tecnologia e de padrões para os próximos anos, pode-se planejar técnicas e estratégias para preservação de registros digitais. Tais técnicas e estratégias são apresentadas na Seção 3.3.

3.3) Técnicas e estratégias para preservação de documentos digitais.

As técnicas e estratégias para preservação de registros digitais têm sido desenvolvidas e testadas em todo o mundo. Qualquer técnica referente à documentos digitais está relacionada ao estado da arte da tecnologia e, portanto deve enfatizar as possibilidades de visualização da informação e não no documento em si. Essas técnicas envolvem processos e mecanismos, que combinam a estrutura lógica do documento e o suporte físico em que está inserido. Nessa seção abordam-se as principais estratégias para a preservação digital e questões relacionadas: *migração, softwares, emulação, encapsulamento, mídias de armazenamento, ferramentas de gerenciamento, metadados*, dentre outras.

A *migração* é considerada em qualquer processo de acesso a longo prazo. Embora possa parecer um processo simples, uma série de cuidados devem ser observados. Especialmente no caso de sucessivas migrações, alterações na formatação do documento podem torná-lo irreconhecível em longo prazo. O processo de migração leva a perda e a nova formatação dos dados, logo é necessário, entender as características das mídias ou suportes digitais de documentos. A transferência da documentação para novas mídias é necessária e prevista para um prazo reduzido, de acordo com o planejamento do setor de TI e não deve ser superior a cinco anos.

A migração consiste na estratégia de substituir o suporte em que os documentos estão armazenados, ou para uma nova versão de *software*, antes que o documento se torne inacessível. Elabora-se um estudo sobre as opções compatíveis com a tecnologia, as inovações e futuras opções, a partir do qual é traçada uma estratégia para os próximos dez anos. Nesse período deverão realizar-se pelo menos duas migrações. Em vista desse período de tempo, escolhem-se as mídias que tenham duração estável e observam-se as condições ideais de armazenamento.

O *software* é o conjunto de programas necessários para preservação e corresponde a um aspecto complexo na análise para migração. Deve-se observar o lançamento de novas versões. No caso do *softwares* proprietários deve-se levar em conta, ainda na fase de planejamento, o custo da aquisição de novas versões e a solidez do fabricante. No caso de *softwares* livres é interessante optar pelos de código aberto e integrar-se nas discussões com atenção a novas versões e padrões.

Além do *software* criador do documento o sistema operacional também deve considerar os mesmos critérios. É importante tentar prever quando a migração será necessária, em geral programas de computador apresentam padrão de mudança. É possível estimar para os próximos dez anos quantas novas versões irão surgir e o custo da compra e migração dos registros. Muitas vezes os documentos são complexos e possuem como anexos documentos multimídias e nesse caso deve se pensar em um conjunto de softwares que formem um pacote necessário tornando a tarefa ainda mais complexa.

O processo de *emulação* propõe a criação de *software* que imite outro mais antigo. A emulação tem se mostrado uma estratégia relevante para preservação uma vez que dispensa a migração de *software* recriando o ambiente do antigo programa dentro do emulador. Esse procedimento evita as perdas do processo de migração, do ponto de vista do *software*, já que o processo de migração para novos suportes continua necessário. A emulação requer um conjunto de metadados intrinsecamente ligados ao documento, pois para que um programa emulador possa ler um arquivo com trinta anos de idade, o responsável pelo programa deve ter todas as informações possíveis.

Deve-se considerar, entretanto que esse processo é bem mais complexo do que aparenta em princípio. Recriar um ambiente obsoleto dentro da complexidade do ambiente informático não é uma tarefa fácil. Deve-se ter um grande cuidado na criação de metadados que forneçam todas as informações necessárias o que é bastante complicado já que não se pode prever o futuro. Além disso como muitos atores tanto de software quanto de hardware estão envolvidos deve-se pensar em cada um deles sob pena de perder-se parte fundamental desse quebra cabeça. Caso opte-se pela emulação deve-se prever revisões periódicas do material a fim de atualizá-lo segundo as tendências tecnológicas.

Existe também uma idéia de manter funcionando *hardware* e *software* necessários a leitura e a interpretação dos registros durante o máximo de tempo possível. Trata-se de uma alternativa problemática, pois conforme lembra Lusenet (2002), “o *software* e o *hardware* tornam-se obsoletos em questão de anos ao invés de décadas”. Por maior que seja o esforço para

manter um computador funcionando, isso será possível em um horizonte temporal limitado. Esse método pode ser usado em caráter emergencial, enquanto não seja produzida uma estratégia mais adequada.

O *encapsulamento* é o processo que consiste em agrupar, dentro de um mesmo suporte, ou dossiê, o conjunto de *softwares* e a descrição das condições necessárias para a reprodução do arquivo ou conjunto de arquivos a serem preservados. É uma técnica interessante para preservação, mas não tão simples. O encapsulamento prevê a aquisição dos documentos segundo critérios bem definidos, a definição dos formatos utilizados para a preservação à longo prazo, criação de um sistema administrador para os arquivos e de políticas que assegurem as atualizações necessárias de *software* e *hardware*.

O encapsulamento é uma alternativa importante para a preservação, pois os documentos digitais, especialmente aqueles da Internet são complexos e possuem conexões com anexos ou *hiperlinks*, tecendo uma intrincada rede de dados. Para que estratégias de preservação sejam bem sucedidas, deve-se levar em conta os elementos que compõem tal rede e criar encapsulamento dos recursos necessários para a visualização do conjunto. Tal técnica é uma das principais ferramentas previstas no sistema *OAIS* para preservação digital .

Na verdade o encapsulamento possui um lugar de destaque dentro das estratégias de preservação. O contexto dos documentos, suas ligações, softwares necessários e metadados podem ser feitos dentro de pacotes de informação. Conforme previsto pelo sistema *OAIS* referência mundial em sistemas de preservação de registros digitais. Esse pacotes podem ser criados a partir de softwares específicos e usados para armazenamento, submissão e acesso a informação, seus metadados e outros requisitos dependentes do estado da arte da tecnologia.

Dentre as estratégias de preservação, cabe ainda citar a importância e a influência das *mídias de armazenamento*, como parte fundamental dentro do processo. Mídias são qualquer suporte para armazenamento de dados digitais. As principais questões relativas a suporte dizem respeito ao local de armazenamento dos dados e ao acondicionamento das mídias. A escolha da mídia é fundamental em uma estratégia deve-se buscar alternativas mais estáveis. A migração entre diferentes tipos de mídias altera metadados, acarreta perdas e modifica a maneira como os dados estão gravados. Assim, uma mudança de mídia é uma tarefa complexa, mas necessária principalmente no caso de aplicação a um grande volume de documentos.

Outra questão em discussão no caso das mídias de armazenamento é a controvérsia sobre o tempo de duração de uma mídia versus o tempo de sua obsolescência tecnológica. Em geral as mídias tendem a se tornar obsoletas antes do comprometimento de sua estrutura física. A

escolha da mídia deve levar em conta o tipo de acesso aos documentos, por exemplo: um documento acessado com pouca frequência pode ser armazenado em mídia móvel como CD-ROM; documentos acessados pela Internet devem ser armazenados em mídias com padrões dinâmicos e compatíveis.

Segundo o *National Archives* (2003) os seguintes critérios técnicos são importantes na seleção da mídia: i) longevidade de aproximadamente dez anos; após esse período provavelmente o suporte estará obsoleto; ii) capacidade compatível com a quantidade de dados a serem armazenados; iii) detecção de erros tanto de leitura e de gravação, com teste da integridade dos dados após a gravação e técnicas de recuperação; iv) obsolescência: utilizar padrões tecnológicos maduros; v) custo: valor por gigabyte, abrangendo o custo da mídia e da propriedade de *software* e *hardware* necessários; vi) susceptibilidade: baixa em relação a danos físicos e as condições ambientais, sem perda de dados.

As *ferramentas de gerenciamento* são fundamentais na pesquisa sobre preservação digital, uma vez que tais ferramentas vão operacionalizar toda a política de preservação da instituição. Em geral, são baseadas em um *software* capaz de controlar o trâmite dos documentos e qualquer ação sobre os mesmos. Um sistema ideal deve levar em conta os preceitos arquivísticos já discutidos, a garantia de autenticidade dos documentos em consonância com os padrões adotados por cada instituição, o controle rigoroso de acesso e quaisquer outras operações realizadas em cada registro. Segundo Hedstrom (1997) a ferramenta deve agregar inteligência sobre o status de preservação dos materiais possibilitando o monitoramento e a manutenção de grandes coleções de objetos digitais de forma automatizada.

Outra questão relevante, de suma importância para as pretensões de preservação de uma instituição são os *metadados*. Metadados tem sido apontados como ferramenta fundamental para a preservação digital e certamente merecem lugar de destaque em qualquer estratégia (LANZINGER, 2001). Na verdade, os metadados são problema e solução para a guarda dos documentos digitais: são responsáveis pela dificuldade em manter documentos digitais na medida em que contém as informações extraordinárias de um registro e ao mesmo tempo, são os descritores responsáveis pelo elo entre o documento e os futuros pesquisadores. Num editor de texto, por exemplo, toda a formatação que vai além do texto é composta por metadados e está sujeita a mudança em cada nova versão ou mesmo em um novo ambiente, como um sistema operacional diferente.

Segundo Ikematu (2001) metadados correspondem a “informação necessária para administrar o recurso *dos dados*”. No caso da preservação digital existe um processo de

administração que engloba uma política institucional de preservação, onde os metadados são peças chaves. Toda a informação que descreve o registro ou conjunto de registros, seu contexto, propriedade intelectual, instanciação e características descritivas estão contidas nos metadados. Todos esses dados sobre o documento são fundamentais para seja possível conduzir migração de *software*, *hardware*, emulação e ainda estar disponível para consulta por um longo tempo.

Lazinger (2001) lembra que metadados são usados para descrever quaisquer dados que tenham como função primária auxílio a identificação, descrição e localização de recursos digitais. São utilizados para recuperação de informação e controle de visualização da informação. Existem metadados criados manualmente e outros automaticamente. Vários padrões foram criados para diferentes aplicações (XML, SGML, AACR2, etc). O mais importante, entretanto é a proposta para a criação de um modelo de descrição eletrônica de dados que facilita construção de sistemas e a manutenção dos registros digitais ao longo do tempo.

Segundo Thomaz (2004), os metadados utilizados na descrição de registros digitais antecedem a Web, mas começaram a se destacar na década de 80 para descrever as características armazenadas em bancos de dados. Os metadados ganham então importância no contexto de preservação digital. São usados no processo de descrição de documentos arquivísticos, tanto físicos quanto digitais, onde funcionam como instrumento de contextualização e identificação.

Documentos digitais estão sempre vinculados a uma formatação básica. O código *ASCII* é um exemplo muito comum. O mesmo possui um código, sempre binário e pré-definido, para cada caractere ou comando básico. Dessa maneira todas as letras, assim como os comando mais básicos executados por um editor de textos são representados por um determinado código. Para além disso cada programa terá uma formatação adicional que será agrupada na forma de metadados. Dentre estes estão os que informam, ao sistema operacional, ou ao conjunto de *hardware* a maneira como estes devem comportar-se para correta interpretação daquele conjunto de dados. Muito embora se possa vislumbrar um futuro com códigos padronizados existe uma grande barreira colocada pelos softwares comerciais que não possuem interesse em compartilhar algo comum, já que customizam seus códigos fontes a fim de obterem vantagens comerciais.

Uma abordagem atual é a tentativa de separação entre o conteúdo do arquivo e sua formatação (metadados), como é o caso do formato proprietário da *Adobe Corporation* (extensão .pdf). A alternativa, entretanto não resolve a questão da preservação, pois além do texto e dos metadados, o *software* de criação precisa de mecanismos para interpretação e para combinação dos dois elementos.

A norma ISO 19005-1 determina a utilização da extensão PDF / A para documentos digitais que necessitem de preservação por um longo período de tempo. O padrão deriva do formato proprietário da *Adobe Corporation em sua versão 1.4* e é uma tentativa de criar um modelo que possa dar suporte interno a metadados em XML, imagens supercomprimidas (a até 20% do TIFF CCITT G4, na mesma resolução), imagens pesquisáveis pelo conteúdo e links de hipertexto. A idéia do arquivo é que todo documento possa ser guardado em um formato de imagem e possa ser recuperado e impresso em qualquer tempo. Embora seja muito interessante o padrão não resolve o problema de documentos mais complexos como sites ou banco de dados que possuem interações complexas na medida em que forma imagens estáticas de documentos.

Cada uma das técnicas descritas são parte do processo de discussão sobre preservação digital. Muitos fatores são importantes nesse contexto e as considerações sobre o assunto vão muito além de simples procedimentos para conservação de dados. Uma vez apresentadas as principais estratégias e métodos para preservação digital, a seção seguinte descreve brevemente as principais iniciativas internacionais sobre preservação digital, com destaque para a questão dos metadados, relevante no contexto do presente trabalho.

3.4) Iniciativas para a preservação digital

A teoria arquivística é aplicada especificamente em cada ambiente e no caso dos documentos digitais, deve-se seguir a mesma regra. A criação de normas satisfaz um conjunto de requisitos funcionais de forma a garantir que o processo de preservação tenha início na produção ou recebimento do documento. Embora o conhecimento dos mecanismos de preservação e modelos gerais seja de auxílio, cada instituição precisa ter o seu próprio modelo inserido em um projeto maior, a política de gestão documental da empresa. Existem padrões específicos para a preservação digital, embora nenhum seja direcionado à realidade brasileira ou mesmo escrito em português.

A mais importante iniciativa para definição de preservação em ambientes digitais é o *OAIS-Open Archives Information System Reference Model*, um padrão ISO para preservação de documentos digitais a longo prazo. Desenvolvido originalmente pelo *CCSDS-Consultative Commite for Space Data Systems* da NASA. Esse comitê se reúne periodicamente para discutir problemas e formular soluções técnicas. Assim, criou-se um documento cujo objetivo é definir a norma ISO para um modelo de referência de preservação de registros digitais à longo prazo.

O modelo foi definido através de uma organização de pessoas e sistemas que aceitou a responsabilidade de preservar a informação digital tornado-a acessível para uma determinada comunidade. É denominado *OAIS-Open Archival Information System* e é aplicável a qualquer

tipo de arquivo digital e especificamente para tornar a informação digital acessível por longo período de tempo. O OAIS é uma fonte muito importante dentro da área de preservação digital por ser base para diversas iniciativas no mundo e possuir uma grande autoridade dentro da comunidade internacional que discute o assunto.

Outros projetos de pesquisa vêm tratando da problemática da preservação digital a partir da prática arquivista. Dentre eles destaca-se o *Projeto Interpares*, baseado no modelo *OAIS-Open Archival Information System*, que pode ser entendido como:

uma estrutura conceitual que disciplina e orienta um sistema encarregado de preservar por longo prazo e manter o acesso a informação digital de qualquer natureza” (*CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEM*, 2002, p.5).

O projeto *Interpares* divide-se em três fases e tem como objetivo desenvolver conhecimento teórico e metodológico para a preservação permanente de arquivos gerados eletronicamente. Com base nesse conhecimento, espera formular modelos de estratégias, políticas e padrões capazes de assegurar a preservação. O projeto está em sua terceira fase e já apresentou modelos para elaboração de políticas para preservação digital que abrangem a captura do objeto digital, sua organização, trâmite, preservação e garantia de autenticidade.

Metadados podem ser criados automaticamente sem a interferência humana, gerando algum tipo de ambigüidade. Na tentativa de minimizar tais problemas existem padrões de metadados como o *DCMI-Dublin Core Metadata Initiative* (2004), que apresenta um conjunto mínimo de atributos para cada termo segundo o padrão *Dublin Core*, além de outros que podem ser adicionados, caso necessário. Existem ainda diversos padrões baseados no DCMI e destinadas a usos particulares segundo a necessidade de cada país. Ainda assim, o padrão atinge apenas determinadas tipos de informações e tem aplicações limitadas: descreve e indexa documentos, mas não apresenta um método que possa ser usado de maneira universal.

Bearman e Sochats (2004) descrevem o documento *Metadata Requirements Projects*, da *University of Pittsburgh*, que trata de especificações para a criação de padrão de metadados baseado em requisitos emergentes, levantados por profissionais tradicionalmente envolvidos com a informação digital. A especificação define propriedades identificadas legalmente, regulamentos e melhores práticas para a sociedade, reunindo assim propriedades fundamentais dos documentos. O documento traz uma série de requisitos funcionais para os elementos de prova dos registros mantidos, assegurando que os mesmos sejam autênticos durante o processo de preservação.

O *OCLC/RLG Working Group on Preservation Metadata*, no documento intitulado “*A Metadata Framework to Support the Preservation of Digital Objects*” de 2002, baseado no

modelo *OAIS-Open Archives Information System*, descreve outro conjunto de metadados com definições de todos os elementos e processos necessários para armazenar e garantir a autenticidade de registros digitais no longo prazo. Basicamente, o documento propõe uma sintaxe de requisitos de *hardware* e termos correspondentes, e para cada um deles, definições em uma classe e relações com campos básicos (nome, origem, definição, objetivo e exemplo).

O *Cedars Project* corresponde uma iniciativa de pesquisa formada entre 1998 e 2002 por um consórcio entre universidades do Reino Unido. O objetivo era sensibilizar a comunidade internacional para a importância da preservação digital à longo prazo. O *Cedars Project* resultou em artigos e conferências, além de trabalho prático operacionalizado por um protótipo baseado no sistema OAIS, empregando a especificação do conjunto de metadados para conservação de recursos digitais. O conjunto de dados que trata sobre memória descritiva, administrativa, técnica e legal descreve a estrutura necessária para a criação de um pacote de informação que organiza um documento digital para preservação. Esse pacote é composto por informações descritivas e informações de contexto, de onde se originam inúmeras outras subdivisões. A partir da criação do vocabulário e das relações entre os termos, o programa levanta uma série de elementos, posteriormente definidos segundo um conjunto de campos mínimos de metadados para a preservação de um pacote de informação.

O *Preservation Metadata for Digital Collections Initiative*, desenvolvido pela *National Library of Australia* (1999), enfatiza a questão dos metadados na gestão de coleções digitais. Destaca que as possibilidades de reconhecer erros, apenas visualizando o objeto digital é menor do que quando é utilizado o suporte tradicional. Muitas vezes só é possível reconhecer e gerenciar objetos a partir dos dados associados e da quantidade de informação envolvida a qual é, via de regra, muito maior.

Esse fato leva a necessidade de sistemas automatizados, tanto quanto possível, baseados na interpretação de metadados. A Biblioteca Nacional da Austrália, atendendo a demanda de empresas daquele país, tem buscado criar parâmetros para metadados que atendam documentos nati-digitais e também aqueles convertidos por digitalização. A exemplo de outros modelos esse também se baseia no *OAIS*: os metadados descrevem processos de seleção dos objetos a serem tratados no processo de preservação, registro da história dos objetos no processo de preservação, definição das estratégias e alternativas viáveis para o processo.

O *METS-Metadata Encoding and Transmission Standards* é uma iniciativa desenvolvida pelo *Network Development and MARC Standards Office* da *U.S. Library of Congress*. Consiste em um padrão de codificação descritiva, administrativa e estrutural de

metadados para bibliotecas digitais. Possibilita a produção de metadados padronizados para codificação e envio em formato XML, podendo ser sub-dividido em até sete elementos de descrição.

Para estudar as principais iniciativas de preservação leva-se em conta os aspectos que envolvem o documento digital. Cada registro é um objeto complexo que além de si mesmo traz outros atributos agregados na forma de metadados. A complexa rede de atributos que possibilitam a visualização do registro deve ser mapeada. Assim, cria-se uma rede envolvendo as facetas necessárias. A complexidade inerente ao processo torna as iniciativas para estudá-los extensas e nem sempre de abordagem simples.

Finalmente, cabe-nos descrever o CIDOC-CRM, uma iniciativa relevante em nível internacional e essencial no contexto desse trabalho. O modelo fornece definições e estrutura formal para descrever os conceitos implícitos e explícitos e seus relacionamentos na documentação do patrimônio cultural. O objetivo do CIDOC-CRM é possibilitar o mapeamento do patrimônio cultural fornecendo um quadro comum e uma semântica para coleta de informações. Destina-se a ser uma linguagem comum para formular requisitos de sistemas de informação e servir como um guia de boas práticas de modelagem conceitual.

O trabalho é fruto do esforço voluntário de várias pessoas. Em 1999 o CIDOC foi apresentada como norma ISO. A *International Standardization Organization* possui autoridade para declarar o que é válido e recomendado em âmbito mundial o que trouxe grande notoriedade a esse padrão. Isso significa que, se valendo dessa autoridade, a equipe do CIDOC fará um esforço para obter o consenso mais amplo com a comunidade internacional. A idéia é que seja um trabalho colaborativo que possa culminar na normalização. O modelo é extensível e haverá uma necessidade contínua de criação de novas áreas, bem como manter o modelo atualizado devido as inevitáveis mudanças no estado da arte das diversas áreas do conhecimento. A própria aplicação prática da norma trará problemas que realimentarão a discussão e possibilitarão adequações.

A estágio terminológico preliminar proposto nesse trabalho está alinhado com os princípios ontológicos do CIDOC-CRM por dois motivos principais: em primeiro lugar, para fomentar a interoperabilidade semântica, o que reconhecido até mesmo pela autoridade da ISSO; em segundo lugar, pelo fato de ser uma iniciativa que não somente possibilita, mas almeja crescer e nesse sentido abre espaço para novas áreas do conhecimento. Optou-se por, utilizando princípios ontológicos do CIDOC-CRM, organizar a terminologia em língua portuguesa sobre o tema tão importante como a Preservação Digital. Espera-se contribuir para que especialistas e

profissionais ligados ao tema possam adequar os sistemas de informação a os requisitos necessários para manutenção de objetos digitais por longo prazo.

3.4) Implicações para a pesquisa

A presente seção aborda as implicações obtidas com a fundamentação teórica sobre preservação digital. Buscou-se reunir literatura para embasar a pesquisa e determinar um norte para a pesquisa. Não houve a intenção de esgotar toda a literatura da área, mas conseguir uma visão diversificada do que está sendo discutido sobre o assunto na comunidade de pesquisa.

Partindo da análise arquivística apresentada na Seção Dois, observa-se a evolução das técnicas e dos princípios promovendo a arquivística à condição de área do conhecimento independente. Os documentos em si também seguiram a mesma tendência. Os suportes vêm evoluindo e acompanhando os avanços tecnológicos, desde as tábuas de argila, passando por pergaminhos e documentos em papel, até os documentos digitais. Embora os suportes sejam diferentes nas diversas épocas, um problema é comum: como preservar em caráter permanente o que é perecível? Essa questão persegue o homem e, se já representava um grande problema nos suportes tradicionais, este se agravou nos arquivos eletrônicos.

O suporte de acondicionamento de informações eletrônicas é, sem dúvida, o mais frágil da história. Corre-se o risco de perder grande parte da história da humanidade, das informações fundamentais para o avanço científico em questão de anos, caso nada seja feito para garantir o acesso futuro a informações importantes. Muda-se o foco da preservação que não é mais o objeto, mas a garantia de acesso dentro de um universo complexo com objetos binários distintos e intrinsecamente ligados através de uma grande teia de dados e metadados. Tais objetos somente podem se tornar informação quando relacionados com outros dados dentro de um contexto específico que muda com uma velocidade crescente.

Através da fundamentação teórica percebeu-se que a Arquivologia, com teorias de base como o princípio da proveniência ou respeito pelos fundos, organicidade e outros fundamentos, pode auxiliar na contextualização dos registros que precisam ser preservados no longo prazo. Além disso, os arquivos, entendidos enquanto instituições administradoras de documentos são guardiões da história da humanidade na medida em que contém as provas que baseiam as pesquisas dos profissionais historiadores. Muitos documentos importantes que testemunham o início da era digital estão definitivamente perdidos por falta de uma política de preservação e isso pode significar atraso à humanidade caso nada seja feito.

Além dos princípios da Arquivologia destacam-se temas abordados sobre preservação digital: i) mídias de armazenamento: assim como a informação é efêmera e muda com rapidez, todo processo de preservação deve levar em conta a migração para novos suportes; ii) emulação: consiste em criar programas ou equipamentos que possam simular as condições onde os registros foram criados; iii) migração: permite a conversão de documentos para um suporte que seja acessível, e está relacionada tanto a *hardware*, quanto a *software*; iv) preservação de *hardware* e *software*: relacionado a emulação, esse procedimento consiste em preservar as máquinas e o programas para o funcionamento dos registros; v) encapsulamento: consiste em agrupar em mesmo suporte o conjunto de *softwares* e a descrição das condições necessárias para a reprodução do arquivo; vi) metadados: definidos de forma simples como dados sobre dados, são a base dos registros digitais e da preservação em longo prazo.

Acredita-se assim, que ao confrontar aspectos da arquivística tradicional (Seção Dois) com as questões da preservação digital no âmbito dos Arquivologia (Seção Três) foram estabelecidas as bases para execução da pesquisa, a qual será descrita na Seção Quatro, a seguir.

4) METODOLOGIA DE PESQUISA

Essa pesquisa se baseia em um estudo interdisciplinar: tomando como fundamento a Arquivologia, seus métodos e conhecimento, e aliando-se a princípios ontológicos objetiva-se apresentar um conjunto de metadados que poderá, no futuro, servir como apoio para projetos de preservação de registros digitais em longo prazo, para a construção de ontologias e para a integração entre sistemas.

A Seção Dois do trabalho apresentou os fundamentos da arquivística subdividindo-se em conceitos básicos, instrumento de controle e organização de arquivos e gerenciamento de documentos de arquivo. A Seção Três descreveu a preservação digital subdividiu-se em preservação digital no âmbito da arquivística, requisitos para a preservação digital, técnicas e estratégias para preservação digital e iniciativas.

A presente Seção descreve as etapas da metodologia de pesquisa. Do ponto de vista da metodologia científica, a proposta pode ser classificada segundo (LAKATOS e MARCONI, 1991) como: pesquisa de natureza aplicada, objetivando criar conhecimento para aplicações de preservação digital; de abordagem qualitativa, uma vez que seus resultados não são mensurados numericamente; de caráter exploratório e experimental, uma vez que aborda uma temática ainda em estudo pela área; e de objetivo explicativo.

O restante da Seção está organizado da seguinte forma: a Seção 4.1 descreve o objeto de pesquisa, uma terminologia de termos representativos de metadados para a preservação digital construída através da análise de outras iniciativas; a Seção 4.2 detalha a metodologia de pesquisa subdividindo-a em quatro etapas.

4.1) O objeto da pesquisa

O objeto de pesquisa é um conjunto de termos representativos de metadados para preservação digital, para a proposição de melhorias a partir do refinamento das relações entre os metadados das diversas iniciativas. De forma, colabora-se na busca por um padrão de metadados para preservação digital, adequado à realidade da arquivística brasileira. A descrição dos termos representativos de metadados e respectivas relações, semanticamente definidas, são representadas em uma estrutura de organização ontológica, baseada na CIDOC-CRM.

Vários padrões de metadados têm sido usados e propostos para preservação digital. As principais iniciativas que apresentam metadados descritivos de objetos digitais segundo (CONSTANTOPOULOS e DRITSOU, 2007), os quais foram avaliados, são:

- *Metadata Requirements Project*, da *School of Library and Information Science, University of Pittsburgh*, Estados Unidos;
- *DCMI-Dublin Core Metadata Initiative*, organização sem fins lucrativos mantida pela *OCLC- Online Computer Library Center Research*;
- *OAIS-Open Archives Information System Reference Model*, um padrão ISO para preservação de documentos digitais a longo prazo, desenvolvido originalmente pelo *CCSDS-Consultative Commite for Space Data Systems* da NASA;
- *Cedars-CURL Exemplars in Digital Archives Project*, da *CURL-Consortium of University Research Libraries*, Reino Unido;
- *Preservation Metadata for Digital Collections Initiative*, da *NLA-National Library of Australia*.
- *METS-Metadata Encoding and Transmission Standards*, um padrão desenvolvido pelo *Network Development and MARC Standards Office* da *Library of Congress*.

Embora não seja voltada especificamente para a preservação de registros digitais a Norma Brasileira de Descrição Arquivística também foi incluída, uma vez que é importante ferramenta de apoio a classificação pós-coordenada em arquivos. Considera-se que essa iniciativa reforça a relação existente entre os princípios arquivísticos e a preservação de registros digitais.

De forma a manter o princípio da interoperabilidade entre estágio terminológico preliminar aqui proposto com a maioria dos padrões internacionais, são adotados princípios ontológicos definidos no *CRM-Conceptual Reference Model* do *CIDOC-International Committee for Documentation*⁷, mantido pelo *ICOM-International Council of Museums*⁸. Tal modelo de referência, conhecido como *CIDOC-CRM*⁹ (CROFT et al, 2001) é a mais completa ontologia formal para integração e intercâmbio de dados sobre patrimônio cultural digital.

A criação da terminologia como termos representativos de metadados para preservação digital permitiu a avaliação das possibilidades de integração das iniciativas existentes, conforme já apontado pela pesquisa internacional (CONSTANTOPOULOS e DRITSOU, 2007; BERMÈS e POUPEAU, 2008; DUBIN e PLUTCHAK, 2006). Para tal, foi avaliada a semântica de cada iniciativa, e uma vez comparadas, os resultados foram integrados

⁷ Disponível em <<http://cipa.icomos.org>>. Access: 20 July 2008.

⁸ Disponível em <<http://icom.museum>>. Access0: 16 oct. 2008.

⁹ Disponível em <http://cidoc.ics.forth.gr>>. Access0: 17 oct. 2008.

ao CIDOC. As Seções 4.2 e 4.3 a seguir apresentam, respectivamente, as etapas da metodologia de pesquisa e os resultados esperados

4.2) Descrição da metodologia de pesquisa

A **primeira etapa** consiste em *avaliar as iniciativas de metadados para preservação digital* selecionadas. Para tanto, as seguintes tarefas foram realizadas:

- Entendimento das propostas subjacentes a cada iniciativa, bem como seu contexto e sua utilização, uma vez que existem iniciativas especificamente para preservação digital e outras para objetos de bibliotecas digitais.
- Criação de tabela contendo:
 - listas de termos representativos dos metadados utilizados em cada iniciativa, bem como as definições apresentadas para cada termo em seu contexto de aplicação;
 - Lista comparativa entre os termos, de forma a verificar de forma preliminar a consistência entre as definições;

A **segunda etapa** consistiu em *descrever a semântica de cada elemento de metadados*, em cada iniciativa selecionada em linguagem natural (português), uma vez que os profissionais da arquivística que lidam com preservação digital, além de outros profissionais (por exemplo, cientistas da computação, bibliotecários, gerentes, dentre outros) necessitam de definições claras, abrangentes e contextualizadas.

Quando necessário, as definições dos elementos das iniciativas foram comparadas, de forma a obter as melhores possibilidades de definição dos termos e relações entre os elementos de metadados. Adota-se como termo oficial aqueles termos considerados obrigatórios em cada iniciativa, bem como aqueles que se repetem em duas ou mais iniciativas. Admite-se ainda a inclusão de alguns termos, uma vez que eles são essenciais dentro do contexto arquivístico ou classes fundamentais para o agrupamento dos elementos.

A **terceira e última etapa** consiste na inserção da terminologia adquirida nas etapas anteriores abaixo da estrutura de alto nível, a CIDOC-CRM, após avaliação da melhor localização do termo de acordo com a caracterização correspondente. O fragmento obtido, uma vez alinhado ao CIDOC-CRM, será passível de integração com outros sistemas de classificação, automáticos ou não.

Para facilitar a visualização da metodologia proposta como um todo, a FIG.4 apresenta o quadro sinótico com breve descrição de cada etapa:

| Etapa | Tarefa | Instrumentos |
|--------------|---|---|
| Etapa 1 | Avaliar conjuntos de metadados utilizados para preservação digital. | Iniciativas internacionais. |
| Etapa 2 | Descrever a semântica dos elementos de metadados. | Iniciativas internacionais, formulários. |
| Etapa 3 | Inserir termos e respectivas definições no CRM-CIDOC. | Iniciativas internacionais, quadros comparativos, quadro comparativo final. |

Figura 4: quadro sinótico de resumo das atividades

5. COLETA DE DADOS

Neste capítulo são abordadas seis iniciativas anteriormente mencionadas com o objetivo de criar uma lista contendo o conjunto de metadados associados a cada iniciativa. As seções de 5.1 até 5.7 apresentam o resultado para cada iniciativa mencionada na metodologia. A seção 5.8 apresenta uma breve consideração sobre as implicações para a pesquisa como um todo.

5.1) *Metadata Requirements Project*

A primeira iniciativa estudada é a da Universidade de *Pittsburgh*, nos Estados Unidos que está trabalhando em um modelo de referência em **comunicação para negócios**. Tal modelo é denominado como *Metadata Requeriments Project* e reúne gerentes de aplicação em domínios de comércio especialmente nos ramos de saúde e desenvolvimento para fábricas. A idéia do projeto é tentar definir padrões de metadados que possam garantir a autenticidade no intercâmbio de dados confiáveis para transações comerciais.

Ao mesmo tempo busca-se identificar os requisitos necessários para: controlar o acesso, gerenciar dependência de *Software*, representar o negócio de forma simplificada, documentar o uso de dados dentro de um ambiente complexo, amplo e heterogêneo como a computação.

O estudo gerou uma especificação que define treze características oriundas de argumentos legais, regulamentos e das melhores práticas em toda a sociedade. Com isso, determinou as propriedades consideradas fundamentais nos registros que serão preservados por um longo período de tempo. Essas características podem ser definidas como “regras de produção” ou “declarações de lógica simples de atributos observáveis”. O grande problema, observado por seus especialistas, é que os requisitos descritos através da linguagem natural são muitas vezes ambíguos e permitem diferentes interpretações ligadas a subjetividade e contexto dado por cada um.

Para minimizar esse problema o grupo definiu um conjunto de regras de produção formais. Este foi utilizado durante o processo de desenvolvimento de um grupo mínimos de exigências e na construção da representação das informações. Isso ajudou a construir processos mais explícitos retirando o máximo possível o caráter subjetivo. Também garantiu uma especificação lógica e mais refinada através da observação de estados ou propriedades. Desde a criação do registro existe um conjunto específico de metadados que, quando presentes, satisfazem as necessidades informacionais da especificação. Esses metadados estão indissolúvelmente ligados ao objeto informacional, nesse caso transações comerciais, garantindo

que o mesmo possa ser acessado por um longo período de tempo segundo regras estabelecidas por seu criador, além de ter propriedades totalmente confiáveis para fins de comprovação de negócios. Metadados adicionais mantidos pelo sistema garantem a autenticidade e, portanto o caráter probatório dos documentos digitais em questão.

No intuito de facilitar a aplicação das “provas eletrônicas” nos diversos ambientes o projeto tem proposto, em suas conclusões, um passo mais a frente propondo um modelo de referência denominado *Business Acceptable Communications (BAC)* composto de metadados que garantem que o documento sirva como evidência (prova) do negócio.

Por tratar de um tipo específico de informação, a comunicação de negócios, o modelo faz uma importante distinção entre “Registros” e “Dados”: os primeiros são aqueles trabalhados dentro de um ambiente que permitam a autenticidade dos documentos de modo que possam ser caracterizados enquanto registros de transação para negócios; no segundo caso, pensa-se em ambientes que apenas apresentem dados sem possuir os requisitos necessários para garantia de prova. O trabalho da Universidade de *Pittsburgh* refere-se apenas ao primeiro caso quando faz referência a elementos de prova para uma transação.

Existe grande dificuldade em criar modelos ou requisitos já que ambientes diversos utilizam *softwares* diferentes e na maioria dos casos não se preocupam com a garantia da autenticidade já que isso, pelo menos aparentemente, não contribui para a eficácia corporativa dentro do dia a dia das empresas.

A Universidade de *Pittsburgh* buscou identificar dentro das fontes legais, melhores práticas e discursos sociais aqueles que tratam especificamente do contexto de prova. Embora o modelo proposto trabalhe mais especificamente com as comunicações de negócios essas informações só serão consideradas como evidência caso o registro que representa o negócio tenha sido capturado e preservado segundo um conjunto de requisitos e metadados previamente definidos e que atendam requisitos sob diversos pontos de vista que definem o modelo de referência.

Existe uma preocupação em garantir a interoperabilidade dos sistemas que enviam e recebem registros. Isso poderia ser garantido através de um conjunto de metadados de identificação que descrevessem o registro antes do mesmo ser aberto, pois não se pode garantir o uso de sistemas iguais, o que seria o ideal. Para que a captura das informações que seja crítica para efeitos comprobatórios e de preservação os metadados devem garantir um encapsulamento dos objetos em camadas criando uma modularidade. Para tanto, o sistema deve criar automaticamente metadados sobre: documentação das modificações, autenticações, acessos, migração e distribuição.

Essa criação simplifica a gestão dos registros em grandes redes e pode, desde que adequado a um padrão de metadados dentro de um padrão, apoiar auditorias de empresas no que diz respeito a *Software*, *hardware* e *registros* e sua utilidade como provas. O modelo reconhece, no entanto que para um sistema de auditoria mais completo seriam necessárias outras redes de metadados. Por isso a proposta é um modelo de referência para a comunicação de negócios que seja aceitável e que tente abordar requisitos específicos ao diálogo entre as partes na transação. Esse modelo deve ser agregado a mecanismos que envolvam os demais requisitos fundamentais através de uma estrutura geral de metadados e encapsulamento de objetos.

O primeiro agrupamento de elementos de modularidade funcional foi organizado em seis camadas (módulos) cada módulo possui um conjunto de metadados que preenchem uma camada. Foram geradas 13 características a serem observadas para a garantia da autenticidade dos registros que caracterizam as transações de negócios em meio eletrônico. A seguir, conforme previsto na metodologia de pesquisa, resumem-se as treze características e a lista de metadados separados por módulos a fim de ter-se uma idéia geral da iniciativa dentro de sua perspectiva prática dentro do contexto descrito até aqui:

Modulo Organizacional:

- **Consciência organizacional:** As organizações devem estar registradas legal e administrativamente dentro da jurisdição além de demonstrarem um conhecimento das melhores práticas para a indústria ou empresas do segmento a que pertencem.

Modulo de Sistemas:

- **Sistema de preservação de registros:** Esses sistemas deverão documentar de maneira rigorosa e formal políticas, responsabilidades e metodologias para sua gestão
- **Execução:** Os Sistemas devem ser executados em todos os momentos no curso de todas as atividades do negócio
- **Consistência:** Os sistemas devem processar a informação de uma forma que assegure que os registros possuam credibilidade

Modulo de Captura de Registros:

- **Abrangência:** Devem abranger todas as transações comerciais; todas as informações trocadas entre as partes devem ser registradas
- **Identificação:** Cada registro, identificado de maneira única, deve estar ligado a uma operação (transação) que se utiliza de um conjunto de registros para

criar um registro principal com a soma de todos os dados associados a transação comercial.

- **Completude:** Deve englobar o conteúdo, estrutura e o contexto da operação. A relação entre os diversos registros deve ser clara de modo a refletir toda a transação comercial. As ligações entre os diversos contextos operacionais da relação comercial entre as partes envolvidas deve estar explícita.
- **Autorização:** Todos os registros devem ser criados por um responsável autorizado e documentado, seja ele uma pessoa ou um setor.

Modulo de Manutenção

- **Preservação:** Devem refletir o conteúdo, estrutura e contexto dos sistemas ao longo do tempo que forem preservados.
 - a. Para tanto precisar ser invioláveis, protegidos contra danos ou destruição acidental e toda e qualquer modificação.
 - b. Dentro de qualquer processo de migração deve-se conservar as condições que remetam a época da criação dentro do *software* original.
 - c. O registro deve ser fiscalizável, para isso deve se criar um histórico que registre qualquer ato relacionado ao arquivo incluindo visualização, reprodução, cópia, envio e qualquer outra atividade. Todo registro deve possuir uma trilha de auditoria que será conservada mesmo depois de sua destruição.
- **Auditável:** Devem-se definir níveis de privilégio para modificação de conteúdos dos registros. Toda modificação deve ser documentada nas trilhas de auditoria. Ainda que os conteúdos sejam removidos os registros da auditoria devem ser mantidos.

Modulo de Utilização

- **Exportação:** Deve ser possível exportar os registros para outros sistemas sem perda de informação. Toda exportação deve ser reversível
- **Acessibilidade:** Os registros devem estar sempre disponíveis; sua representação deve sempre ser igual àquela criada no ambiente original para que tenham valor probatório.
- **Controle de Acesso a conteúdo:** Deve haver controle de cópias e de censura de partes que sejam sigilosas e no caso de liberação das partes não públicas deve haver registro nas trilhas de auditoria.

A idéia é que esse modelo coloque os requisitos funcionais de prova no centro de qualquer discussão sobre a aceitabilidade de uma comunicação para negócios. O modelo ressalta que existem muitos outros requisitos, além da prova, fundamentais para um ambiente de rede, contudo prioriza a questão probatória em sua análise.

Não se pode abrir mão das particularidades de cada organização e das pesquisas necessárias para definição de padrões segundo cada caso. Investigações continuarão a ser necessárias para determinar a melhor forma de proporcionar acesso aos registros de tipos específicos de transações, além dos requisitos para migração e controle das versões e ruídos produzidos nesse processo. A abordagem proposta nesse caso não inclui, dentro dos registros a serem preservados dados sobre migração e nem backup por considerar que isso já é parte integrante do dia a dia das organizações.

Nota-se uma forte tendência para o elemento de prova como centro das discussões e uma separação clara dos documentos que devem ser preservados na referência do modelo a “registros” e não a “dados”. Essa separação remete a noção de documento arquivístico anteriormente citada nesse trabalho e a necessidade do estabelecimento de uma política de informação que determine, através de instrumentos de arquivo, a relevância de cada documento segundo o contexto em que se encontra.

No anexo 01 apresenta-se a lista de Termos representativos dos metadados, referente a iniciativa em epígrafe foi disponibilizada no anexo 01 desse trabalho. Estes são divididos em seis camadas que englobam os treze pontos descritos anteriormente:

5.2) DCMI-Dublin Core Metadata Initiative

A segunda iniciativa avaliada é a: *DCMI-Dublin Core Metadata Initiative*, organização sem fins lucrativos mantida pela *OCLC- Online Computer Library Center Research*. A organização é aberta e se empenha no desenvolvimento de padrões de metadados interoperáveis online que suportem uma vasta gama de modelos de negócios. Incluem trabalhos sobre arquitetura e modelagem, comunidades DCMI e grupos de tarefas além de conferências e Workshops.

O DCMI prevê normas simples para facilitar a descoberta, partilha e gestão da informação através de:

- Desenvolvimento e manutenção de padrões internacionais para a descrição de recursos;
- Apoio a comunidade internacional de usuários e desenvolvedores;

- Promoção da utilização generalizada de soluções Dublin Core;

As principais características de uma organização DCMI são os três “I”s:

- **Independente:** não é controlada por interesses comerciais específicos ou outros interesses de domínios específicos ou soluções técnicas específicas
- **Internacional:** DCMI incentiva a participação de organizações em todo o mundo, respeitando as diferenças culturais e lingüísticas
- **Influenciável:** DCMI é uma organização aberta que objetiva o consenso entre seus participantes, não existem pré requisitos para participação.

Introdução e Definições:

Termos Básicos:

- **Nome (Name):** Um símbolo atribuído ao termo deve ser único dentro do contexto DCMI
- **Label (Etiqueta):** Um rótulo, legível por linguagem humana, atribuído ao termo;
- **URI (Uniform Resource Identifier):** Um identificador de cada termo;
- **Definição (Definition):** Uma declaração que representa a natureza e o conceito essencial do termo;
- **Tipo do Termo (Type of Term):** O Tipo de Expressão, tal como descrito no DCMI Abstract Model (DCAM)*

Caso seja necessário os termos podem ser descritos com informações adicionais:

- **Comentário (Comment):** Informação Adicional sobre o termo ou sua aplicação;
- **Veja (See):** Documentação autorizada relacionada com o termo;
- **Referências (References):** Recurso que sejam referenciados na definição ou comentários;
- **Refinamento (Refine):** Uma propriedade que descreve uma sub-propriedade do termo;
- **Mais Amplo (Broder than):** Refere-se a classe do termo, nesse caso descrita como uma super classe;
- **Mais Estreito (Narrower Than):** Refere-se a classe do termo, nesse caso descrita como uma sub classe;
- **Domínio (Hás Domain):** Refere-se a um recurso descrito em determinada instância (dentro de certo domínio do conhecimento, por exemplo);
- **Faixa de Valor (Hás Range):** Refere-se a valores possíveis dentro de determinada instância para aquele termo, como uma máscara.

- **Membro de (Member of):** Determina o esquema de vocabulário a que o termo pertence, que é membro;
- **Instância de (Instance of):** Determina a instância (Classe) a que o termo pertence;
- **Versão (Version):** Um histórico específico de descrição do termo.

Além das definições básicas do Padrão Dublin Core são apresentadas mais sete tabelas de contexto de aplicação para esses metadados. Em todas as descrições são utilizados os termos básicos descritos na introdução e em alguns casos acrescentam-se os termos adicionais, também previamente descritos:

- Propriedades gerais dos termos: metadados que descrevem as propriedades gerais dos termos ou da instância em que estão inseridos.

| <i>Nome do Termo: resumo</i> | |
|------------------------------|---|
| <i>LPR</i> | http://purl.org/dc/terms/abstract |
| <i>Etiqueta:</i> | Resumo |
| <i>Definição:</i> | Um resumo do recurso |
| <i>Tipo de termo:</i> | Propriedade |
| <i>Refino:</i> | http://purl.org/dc/elements/1.1/description |
| <i>Refino:</i> | http://purl.org/dc/terms/description |
| <i>Versão:</i> | http://dublincore.org/usage/terms/history/#abstract-003 |

Figura 06: Exemplo de descrição do termo resumo do padrão *Dublin Core*

- Propriedade de Legado: descrevem a abrangência, e origem da criação do termo

| <i>Nome do Termo: Data</i> | |
|----------------------------|---|
| <i>LPR</i> | http://purl.org/dc/elements/1.1/date |
| <i>Etiqueta:</i> | <i>Date</i> |
| <i>Definição:</i> | <i>Um ponto ou período de tempo associado a um evento no ciclo de vida de um recurso</i> |
| <i>Comentário:</i> | <i>Data que será usada para expressar uma informação temporal em qualquer nível ou granularidade.</i> |
| <i>Referências:</i> | [W3CDTF] http://www.w3.org/TR/NOTE-datetime |
| <i>Tipo do Termo:</i> | Propriedade |
| <i>Versão:</i> | http://dublincore.org/usage/terms/history/#date-006 |

| | |
|-------------|--|
| <i>Nota</i> | <i>(http://dublincore.org/documents/dcmi-terms/) for an explanation. Um segunda propriedade que ocorre quando o nome da propriedade foi declarado nos termos do Dublin Core.</i> |
|-------------|--|

Figura 07: Exemplo de termos relacionado com o legado no padrão Dublin Core

- Esquemas de codificação de vocabulário: definem esquemas de classificação envolvidos no vocabulário ou necessários para a compreensão dos metadados:

| Nome do Termo: CDD | |
|---------------------------|--|
| <i>LPR:</i> | <u>http://purl.org/dc/terms/DDC</u> |
| <i>Etiqueta:</i> | <i>DDC</i> |
| <i>Definição:</i> | <i>O conjunto de recursos conceituais especificados no sistema e classificação decimal.</i> |
| <i>Veja:</i> | <u>http://www.oclc.org/dewey/</u> |
| <i>Tipo de Termo:</i> | <u>http://purl.org/dc/dcam/VocabularyEncodingScheme</u> |
| <i>Versão:</i> | <u>http://dublincore.org/usage/terms/history/#DDC-003</u> |

Figura 08: Exemplo de descrição de esquemas de codificação de vocabulário do padrão *Dublin Core*

- Sintaxe de esquemas de classificação: Apresentam a descrição de esquemas de classificação ou codificação utilizados nas descrições, como por exemplo normas internacionais ou codificações criadas pelo próprio comitê responsável pela atualização do Dublin Core.

| Nome do Termo: ISO3166 | |
|-------------------------------|--|
| <i>LPR:</i> | <u>http://purl.org/dc/terms/ISO3166</u> |
| <i>Etiqueta:</i> | <i>ISO 3166</i> |
| <i>Definição:</i> | <i>The set of codes listed in ISO 3166-1 for the representation of names of countries.</i> |
| <i>Veja:</i> | <u>http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html</u> |
| <i>Tipo de Termo:</i> | <i><u>Datatype</u></i> |
| <i>Versão:</i> | <u>http://dublincore.org/usage/terms/history/#ISO3166-004</u> |

Figura 09: Exemplo de sintaxe de esquemas de classificação de vocabulário do padrão *Dublin Core*

- **Classes:** Representam classes de recursos ou de agrupamento de termos que precisam ser definidos dentro do processo

| <i>Nome do Termo: Frequencia</i> | |
|----------------------------------|---|
| <i>LPR:</i> | http://purl.org/dc/terms/Frequency |
| <i>Etiqueta:</i> | <i>Frequency</i> |
| <i>Definição:</i> | <i>A rate at which something recurs.</i> |
| <i>Tipo de Termo::</i> | <u><i>Class</i></u> |
| <i>Versão:</i> | http://dublincore.org/usage/terms/history/#Frequency-001 |

Figura 10 - Exemplo de classe do padrão *Dublin Core*

- **Vocabulário de Tipos DCMI:** Define tipologias tanto para uso de padrões diferentes de arquivos quanto para coletivos de dados definidos em padrões codificáveis.

| <i>Nome do termo: Imagens em movimento</i> | |
|--|---|
| <i>LPR:</i> | http://purl.org/dc/dcmitype/MovingImage |
| <i>Etiqueta:</i> | <i>Imagem em movimento</i> |
| <i>Definição:</i> | <i>Uma série de representações visuais partindo da impressão de um filme mostradas em sucessão.</i> |
| <i>Comentário:</i> | <i>Exemplos incluem animação, filmes, programas de televisão, vídeos ou um saída visual de uma simulação</i> |
| <i>Tipo de termo:</i> | <i>Classe</i> |
| <i>Mais estreito::</i> | http://purl.org/dc/dcmitype/Image |
| <i>Membro de:</i> | http://purl.org/dc/terms/DCMIType |
| <i>Versão:</i> | http://dublincore.org/usage/terms/history/#MovingImage-003 |

Figura 11: Exemplo de vocabulário de Tipos do padrão *Dublin Core*

- **Termos Relacionados com o modelo abstrato DCMI:** São verdadeiros metadados de metadados. Agrupam os termos criados pelo próprio padrão *Dublin Core*.

| <i>Nome do Termo: Esquema de vocabulário</i> | |
|--|---|
| <i>LPR:</i> | http://purl.org/dc/dcam/VocabularyEncodingScheme |

| | |
|-----------------------|---|
| <i>Etiqueta:</i> | <i>Esquema de Vocabulário</i> |
| <i>Definição</i> | <i>Uma seleção numerada de recursos</i> |
| <i>Veja:</i> | http://dublincore.org/documents/2007/06/04/abstract-model/ |
| <i>Tipo de termo:</i> | <i>Classe</i> |
| <i>Versão:</i> | http://dublincore.org/usage/terms/history/#VocabularyEncodingScheme-001 |

Figura 12: Exemplo de vocabulário de Tipos do padrão *Dublin Core*

5.3) OAIS-Open Archives Information System Reference

OAIS-Open Archives Information System Reference Model, um padrão ISO para preservação de documentos digitais a longo prazo, desenvolvido originalmente pelo CCSDS-Consultative Commite for Space Data Systems da NASA. O documento emitido pelo comitê CCSDS é uma recomendação técnica desenvolvida por um amplo consenso sobre o que é necessário para manter-se uma informação digital por um longo período de tempo ou em caráter permanente.

O objetivo do documento produzido foi a criação de uma norma da *International Standarization Organization* (ISO) que trata da preservação de registros digitais a longo prazo. Tal norma foi denominada OAIS (*Open archival information System*). Considera-se longo prazo qualquer tempo suficiente para causar preocupação com a evolução das tecnologias sejam elas novas mídias, formatos de dados ou mudanças no usuário ou comunidade.

Para um entendimento adequado do sistema de arquivo OAIS é fundamental uma compreensão dos conceitos de dado, informação e conhecimento dentro da perspectiva abordada. Um dado, em formato digital é armazenado com uma seqüência que denominamos *Bit Stream*. A codificação desses dados é interpretada por esquemas de representação da informação e a soma desse conjunto forma um objeto informacional inteligível pelo usuário final.

A preservação desse objeto a longo prazo passa pela compreensão do próprio objeto e seus metadados, ou seja compreender os *bits* e a representação ou codificação que se aplica a esses *bits*. Essa tarefa não é fácil, na verdade é mais complexa do que aparente em princípio. Além dos dados e dos padrões deve-se levar em conta que o objeto essencial do OAIS é preservar informação para determinada comunidade em determinado contexto e nesse sentido deve-se levar em conta a base de conhecimentos inerente a esse processo. Um exemplo claro disso são os conceitos e processos usados dentro de instituições que levam em conta a prática e as diretrizes necessárias para o desenvolvimento das atividades além de costumes que podem ser regionais. Esses conceitos mudam com o tempo e precisam ser atualizados juntamente com os

processos de preservação para que, no futuro, se possa contextualizar os dados transformando-os em informação. Processos de negócios, descritos na forma de dossiês, por exemplo, podem sofrer alterações a qualquer momento e após alguns anos serem estranhos ao conhecimento dos funcionários de certa organização, tornando os processos antigos inteligíveis, ainda que tenham sofrido um processo de migração que possibilite sua visualização.

Outro fator possível é a mudança da comunidade em questão o que requer a revisão dos conceitos de contextualização. Um estudo científico que após a sua conclusão passa a ser público ou certo arquivo que deixa de ser protegido pela legislação em decorrência do tempo e por isso passa a ser disponibilizado para o público geral, são casos que requerem uma revisão dos metadados a fim de tornar o arquivo acessível para um novo tipo de consumidor de informação. Deve-se observar se a informação requer a interpretação de especialistas e de padrões internacionais que possam ser alterados ao longo do tempo. Daqui a cem anos nem todas as profissões existentes hoje ou os padrões de medição e registro de dados serão os mesmos, logo se deve tentar controlar essas alterações, através da identificação dos riscos que podem tornar a informação irreconhecível no futuro.

A informação precisa, essencialmente dos dados e dos metadados para sua preservação assim é conveniente definir um conceito de **pacote de informação**. Esse pacote é um *container* conceitual com dois tipos de informações: informações sobre o **conteúdo da informação** (*content information*) e **informação de Descrição para a Preservação** (*Preservation Description Information*) que se denominam aqui metadados, conforme descrito na FIG. 14 extraída do modelo OAIS.

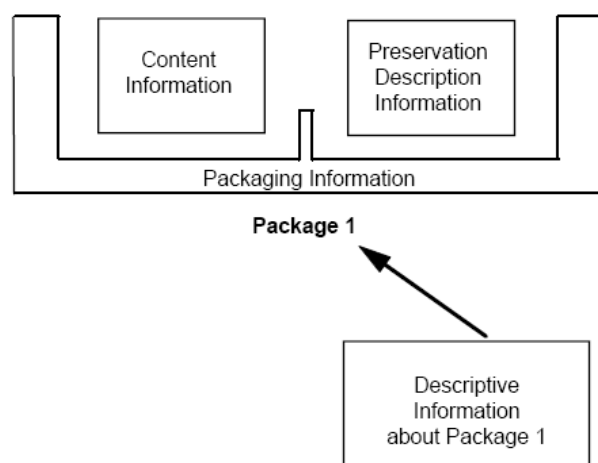


Figura 13: Representação gráfica do pacote de informação proposto pelo modelo OAIS

O OAIS define quatro campos de informações fundamentais para preservação de pacotes de informação:

- Proveniência: descreve a fonte do conteúdo de informação que teve custódia desde sua origem incluindo toda sua história e processamento;
- Contexto: descreve a forma como o conteúdo informacional daquele pacote se relaciona com o ambiente em que ele se encontra, a forma como foi criado e dentro de que processo;
- Referência: fornece um ou mais identificadores a cada objeto de maneira que esse possa ter uma identidade única, semelhante ao ISBN (*International Standard Book Number*).
- Inviolabilidade: fornece um *wrapper* (“um envelope”) que impede qualquer tipo de alteração no pacote.

O OAIS representa um intermediador, um guardião do pacote de informações que além de preservar vai disseminar informações recebendo-as de um produtor, e fornecendo-as, com garantia de autenticidade, para um consumidor, conforme descrito na FIG.15 abaixo:

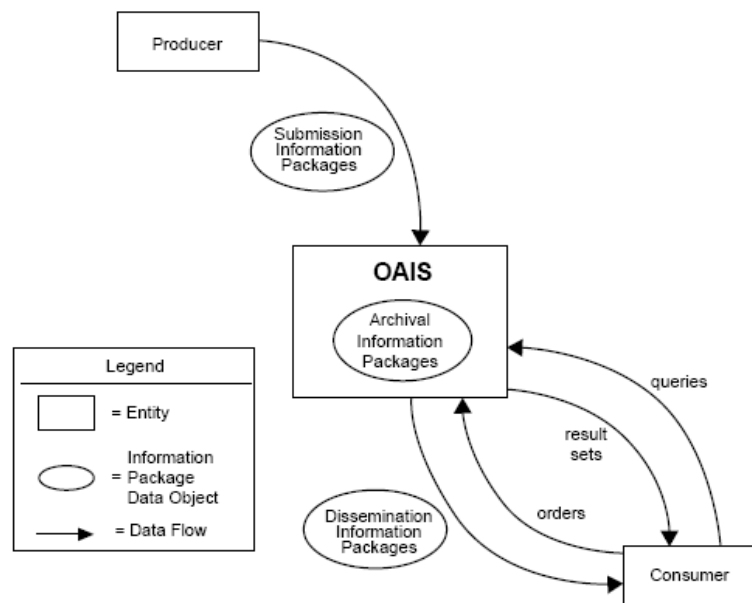


Figura 14 – Esquema OAIS de arquivamento e disseminação de dados

Essa interação entre as partes e especialmente seus limites devem ser definidos através de políticas de informação que são determinadas como **Gestão da Interação**. A gestão vai determinar as relações entre o OAIS e os produtores (no envio de informações para preservação) e os consumidores (quais consumidores e em que condições terão acesso a essas informações). Devem ser definidas ainda as responsabilidades do OAIS em relação a questões inerentes à informação como níveis de acesso, privilégios e sigilo. Embora o OAIS deva possuir liberdade para executar todos os mecanismos de

preservação isso não significa que tenha a posse da informação, mas apenas a custódia da mesma, a propriedade será sempre atribuída ao produtor.

O modelo OAIS apresenta seis entidades funcionais e interfaces relacionadas. Vemos novamente que o OAIS funciona como um intermediador da informação e que além de preservar deve administrá-la.

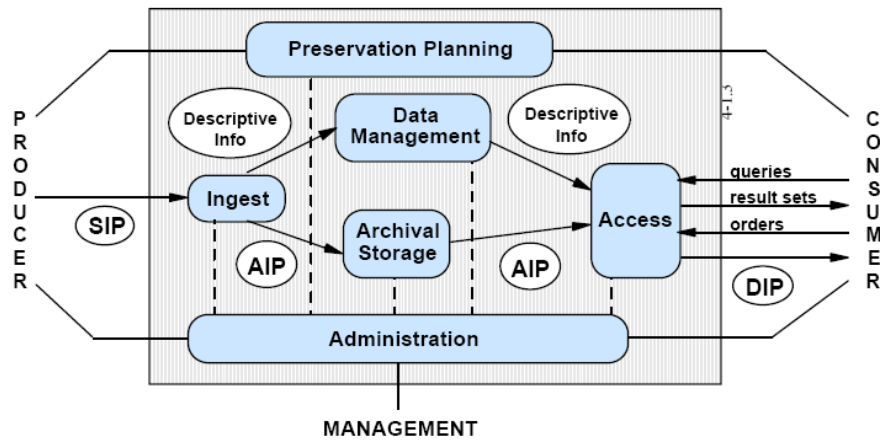


Figura 15: Entidades Funcionais OAIS (Open Archival Information System)

Percebe-se que o OAIS recebe a informação de um produtor criando informações descritivas que somadas aos arquivos de informação vão formar um pacote de informação (AIP – *Archival Information Package*). A partir disso é realizado um gerenciamento de dados e a guarda dos arquivos propriamente ditos. Do outro lado do esquema estão os consumidores de informação que formulam buscas obtendo resultados.

Percebe-se ainda que toda iniciativa para preservação digital de dados deve passar, necessariamente, por um processo anterior que defina políticas e procedimentos de informação dentro de determinada comunidade. As políticas vão definir as relações entre produtores, administradores e consumidores de informação em todos os níveis. Os procedimentos vão definir os processos e os requisitos de arquivamento para os documentos digitais, quais serão repassados para a administração (OAIS), de que forma e em que condições poderão ser acessados.

Um sistema OAIS deve definir um PDI – *Preservation Description Information*. Esse PDI deve conter informações adequadas para a preservação da informação e de seus metadados, fundamentais para o acesso futuro.

Além dos quatro campos fundamentais para a preservação descritos pela iniciativa é preciso entender que para cada tipo de contexto informacional esses quatro campos apresentar-se-ão de maneira diferente, formando um PDI diferente, conforme ilustrado no quadro abaixo:

| Contexto da informação | Referência | Proveniência | Contexto | Aplicável |
|----------------------------------|--|---|--|--|
| Dados Espaciais | Identificador de Objeto Referencia de periódicos Missão, instrumento, título, atributo | Descrição do instrumento; História de processamento; Instrumento; Modo do instrumento; Mapa da documentação; Especificação da interface de software. | História da Calibração; Dados Relacionados; Missão; História da Fundação. | Sistemas de recuperação da informação |
| Coleções de bibliotecas Digitais | Descrição Bibliográfica; Identificador permanente | Para Coleções Digitalizadas: i) Metadados sobre o processo de digitalização; ii) Caminho para a versão original Para coleções nato digitais: i)Caminho para o documento original; Metadados sobre o processo de preservação: i) caminho para a primeira versão do item da coleção; ii) história das migrações | Caminho para os documentos relacionados no ambiente original; | Assinatura Digital Indicador de autenticidade Dados de Checagem (<i>checksum</i>); |
| Pacotes de <i>software</i> | Nome Autor / Organizador Número da Versão Número Serial | História das Revisões; Detentor da licença; Registros; Direitos autorais; | Arquivos de Ajuda; Guia do usuário; Software Relacionado; Linguagem; | Certificado; Dados de Checagem (<i>checksum</i>); |

Figura 16: Exemplos de Tipos de PDI. Extraído da Iniciativa OAIIS(*Open Archival Information System*)

A iniciativa traz ao todo seis conjuntos denominados entidades funcionais que compõem o sistema OAIIS. Essas entidades possuem um fluxo de dados e metadados considerados básicos e fundamentais para a criação de um sistema de preservação de informação a longo prazo.

5.4) *Cedars-CURL Exemplars in Digital Archives Project*

O projeto *Cedars* começou em 1998 e terminou em 2002, no *CURL-Consortium of University Research Libraries*, Reino Unido e teve como principal objetivo explorar a preservação digital. A idéia é criar uma descrição dos objetos que garanta seu acesso em um período longo de tempo. O objetivo principal do projeto foi: abordar questões estratégicas, metodológicas e práticas além de fornecer orientações em melhores práticas para a preservação digital.

O Projeto *Cedars* dividiu-se, basicamente, em duas fases. Nos três primeiros anos buscou produzir material escrito no formato de artigos ou monografias e num segundo momento, no último ano, buscou divulgar esse material em cinco grandes tópicos:

- *IPR Intellectual Property Rights* (DPI) direito de propriedade Intelectual: Explica: porque questões envolvendo o autor do documento são importantes para bibliotecários e arquivistas que trabalham com recursos digitais; a necessidade de desenvolver políticas coerentes com os direitos autorais como parte das estratégias de preservação digital. Os mais recentes e relevantes recursos jurídicos para bibliotecários e arquivistas. Ressalva-se que nenhuma parte do documento constituí-se de aconselhamento formal jurídico. Muitos conceitos aplicáveis a materiais físicos não abarcam a totalidade dos recursos digitais. Especialmente o conceito de propriedade física que parece estar sendo substituído pelo direito a acesso a informação. Os direitos de propriedade intelectuais em recursos digitais podem nem sempre ser claros e isso pode atrasar ou até mesmo impedir um processo de preservação.
- *Preservation Metadata* (metadados para preservação): destina-se a fornecer uma breve introdução das iniciativas atuais e seus respectivos desenvolvimentos. Visa atender aos interesses das pessoas que tem responsabilidade de preservação digital no Reino Unido, mas pode servir como base para outras iniciativas pelo mundo. O documento visa proporcionar uma lógica para a criação e manutenção dos metadados de preservação que possa apoiar as estratégias de preservação digital como migração ou emulação, uma introdução aos conceitos e terminologias utilizadas em normas ISO e no modelo OAIS; uma breve introdução sobre o Cedros (parte do projeto Cedars responsável pela preservação digital), um esboço sobre metadados de preservação e alguns resultados de outras iniciativas correspondentes e algumas notas sobre a preservação de metadados. A parte mais importante no âmbito desse trabalho são as especificações de metadados incluídas no final da apresentação da iniciativa.
- *Digital Collection Management* (Gerenciamento de coleções Digitais): o objetivo do documento é prover uma visão sobre o gerenciamento de coleções digitais sob o ponto de vista da preservação digital. A gestão de coleções digitais engloba não somente a aquisição, mas a manutenção e o descarte dos documentos. O primeiro passo para qualquer organização é definir suas responsabilidades. A necessidade de preservar varia de uma instituição para outra. Uma biblioteca que recebe depósito legal, por exemplo, deve preservar praticamente tudo, outra que apenas presta apoio pedagógico terá muito menos o que guardar. Deve-se lembrar que a preservação de recursos físicos sempre teve

destaque nas bibliotecas e arquivos. Diversas técnicas foram utilizadas para fazer com que livros e papéis durassem o máximo de tempo possível, contudo existem diferenças básicas entre a preservação tradicional e a digital. Além do fato de preservar-se o acesso e não somente o documento sempre haverá uma rede de recursos que dependem uns dos outros como hardware e software somados a natureza transitória da tecnologia. O gerenciamento é especialmente importante dentro do contexto da preservação digital, pois é necessário que todo o ciclo de vida do documento leve em conta sua guarda por um longo período de tempo afim de que esse propósito não seja inviabilizado no decorrer do caminho. Esse gerenciamento deve entender bem os atores envolvidos no processo para evitar problemas futuros como a necessidade de um software proprietário utilizado por um editor no momento da criação que ocorreu há muito anos atrás e que atualmente não está mais disponível. O gerenciamento das coleções digitais está inserido dentro de uma complexa rede de causa e efeito que deve ser estudada para não provocar surpresas a definição de um plano que envolva uma matriz de riscos e *stakeholders* torna-se fundamental para a definição dessa complexa rede. Deve-se levar em conta, desde a produção do material questões como:

- qual a natureza do material e quais as implicações da mesma para a preservação?
- O que preciso fazer para fornecer acesso ao material a quem precisa?
- Quais ações são necessárias e como e quando devem ser tomadas?
- Porque se deve preservar esse material e para quem?

Essas e outras questões vão levar a definição dos requisitos de gerenciamento e das políticas necessárias para isso. Isso certamente irá incluir: seleção, aquisição, organização, Armazenamento, acesso, re-seleção (para efeitos de preservação) e preservação. Nota-se enfim que a preservação é a ponta final de um processo que se inicia com a seleção do acervo e está intimamente ligado a diversas outras políticas criando uma rede complexa e não pode nunca ser considerada como uma atividade isolada. Problemas na preservação podem inviabilizar a aquisição de determinada coleção, por exemplo. Os requisitos para a preservação digital serão garantidos através de metadados, sem esses a tarefa torna-se muito difícil, senão impossível. A detalhada teia da preservação digital remete a uma visão mais detalhada e voltada para a preservação nas fases de seleção e aquisição de material o que nunca havia sido pensando no caso de materiais não digitais. No momento da definição de políticas de preservação permanente, algumas diretrizes são definidas pelo *Cedars*:

- A forma como a hierarquia de armazenamento é gerida e por quem

- A forma como os suportes de armazenamento são atualizados e com que frequência.
- Como os objetos devem ser divulgados a partir do repositório

O projeto sugere que tanto a estratégia de migração quanto a de emulação são viáveis para a manutenção dos dados, no segundo caso a que se levar em conta o conceito de pacote de dados utilizado pela iniciativa OAIS.

- *Digital Preservation Strategies* (Estratégias para a preservação digital) – A idéia do documento é discutir e apresentar diversos aspectos que envolvem as estratégias de preservação digital sob a ótica do projeto Cedars. Quando se fala em preservação no meio digital fala-se em preservação do acesso e para isso exige que se agreguem mais dados, na forma de metadados para que o objeto possa ser entendido no futuro. Existem os metadados técnicos que se preocupam com a capacidade de acessar o conteúdo do arquivo após a plataforma em que o mesmo foi desenvolvido não estar mais disponível.
 - Os metadados de preservação que se preocupam com a proveniência, descrição, como autoria, local de publicação e data assim como direitos de propriedade intelectual. Uma das propriedades dos objetos digitais é sua complexidade. Uma página Web, por exemplo, contém muito links que estão ligados dentro de vários arquivos dentro de uma árvore de diretórios, para a preservação de todo o conteúdo o projeto Cedars recomenda que toda a estrutura seja convertida em um único arquivo, formando assim um único *bitstream* isso pode ser feito facilmente através de um programa compactador, por exemplo. Ao reverter esse mapeamento pode-se ter problemas com sistemas diferentes, mas os detalhes que possam surgir poderão ser revertidos pela tecnologia futura desde que seja mantida a estrutura e os metadados estejam adequados.
 - O projeto recomenda a adoção de padrões para nomes dos documentos. As estratégias para preservação não estão preocupadas com os metadados de contexto, que são abordados mais claramente dentro do documento sobre metadados, mas sim com os metadados que permitem acesso ao conteúdo intelectual do recurso preservado. Para tanto é necessário registrar a forma como os dados originais (formato de dados) foram mapeados em um *bitstream* e de preferência proporcionar o acesso a um *software* que possa reverter o processo. Para finalizar o quadro técnico deve-se registrar

informações que permitam o acesso intelectual ao documento já que a forma de armazenamento foi criada dentro de uma plataforma computacional que reflete o estado da arte da tecnologia atual. Deve-se ter uma trilha de auditoria que controle a relação entre as versões.

- A cada nova migração surgem novos processos, novos pacotes de informações e logo uma nova versão. Os metadados devem descrever não somente cada versão, mas o processo de passagem de uma para a outra. A versão um deve conter os detalhes do processo de migração para a dois e essa por sua vez deve conter os dados necessários para o acesso a versão anterior caso seja necessária um acesso retroativo. Um bom lema é nunca realizar uma atualização que não possa ser revertida em todos os casos recomenda-se que o original seja sempre mantido como forma de proteção contra problemas de migração. *Cedars* trabalha com uma plataforma que não modifica os formatos dos dados. A idéia é manter o arquivo original e criar um pacote de informação que além dos metadados inclua um emulador. Sempre que os dados migram existe a necessidade de estudar se a forma usada para manter o acesso é adequada e se necessário criar um novo emulador. Sempre que um novo pacote de informações for criado um novo emulador será parte do mesmo de forma a manter o acesso permanente ao arquivo original.

A mesma lógica, da preservação do objeto digital, pode ser utilizada para a conversão de formatos ou para outras estratégias semelhantes o que realmente importa é que a pedra angular da estratégia *Cedars* consiste em manter os arquivos originais em um pacote de informações que contenha além dos metadados de contexto um sistema que seja capaz de recriar as condições em que o arquivo foi gerado.

É fundamental que se identifique os conteúdos significativos dentro de cada objeto, por exemplo, dentro de um site de um jornal temos texto, figura, links e a relação entre eles. Se essa relação não estiver muito clara pode resultar em uma futura recuperação inadequada como um mapa desfigurado ou um monte de palavras soltas, sem uma ordem que indique seu sentido original. Quanto mais complexo for o objeto maior será o esforço para deduzir suas propriedades fundamentais, mas a expectativa é que possa se criar classes de objetos e a partir disso se reaproveitar a análise para os novos documentos.

- *Distributed Archive Prototype* (Protótipos de arquivos distribuídos) – A iniciativa criou um protótipo de um sistema para teste da metodologia e para demonstrar a

viabilidade da preservação digital em longo prazo a partir do sistema Cedars. Esse protótipo possui três grandes objetivos: Manter os recursos armazenados ao longo do prazo; assegurar que o objeto preservado será encontrado; facilitar a compreensão do recurso. Entende-se que a descrição minuciosa do protótipo, embora parte da iniciativa, foge do foco desse trabalho por isso optou-se por não pormenorizar a descrição desse tópico.

- *Structure of the Cedars metadata specification* (estrutura da especificação de metadados do CEDARS) apresentada a seguir no original (em inglês):

Information Package

Preservation Description Information

Reference Information

Resource Description

Existing Metadata

Existing Records

Context Information

Related Information Objects

Provenance Information

History of Origin

Reason for Creation

Custody History

Change History Before Archiving

Original Technical Environments

Prerequisites

Procedures

Documentation

Reason for Preservation

Management Information

Ingest Process History

Administration History

Action History

Policy History

Rights Management

Negotiation History

Rights Information

Copyright Statement

Name of Publisher

Date of Publication

Place of Publication

Rights Warning

Contacts or Rights Holder

Actors

Actions

Permitted by Statute

Legislation Text Pointer

Permitted by License

License Text Pointer

Fixity Information

Authentication Indicator

Content Information

Representation Information

Data Object

5.5) *Preservation Metadata for Digital Collections Initiative*

A iniciativa da Biblioteca Nacional da Austrália tem foco voltado para os metadados. Ressalta-se que a idéia é conseguir a definição de um conjunto de metadados que possam garantir a preservação de longo prazo de diversos tipos de materiais tanto criados em meio digital quanto posteriormente digitalizados. O projeto de metadados levanta uma série de campos onde a preservação torna-se interessante:

- Preservação de coleções e documentação de bibliotecas;
- Gerenciamento de arquivos de publicações digitais on-line, formatos físicos para publicações digitais e coleções digitais ou analógicas de áudio;
- Gerenciamento de projetos de digitalização para coleções de textos e imagens;
- Desenvolvimento de modelos lógicos de dados para implementação de arquivos digitais específicos;
- Base de dados para designer de WEB Sites.

Embora a *National Library os Austrália* – NLA tenha acumulado grande experiência no assunto e desenvolvido com segurança o modelo em questão os próprios gestores manifestam o desejo de manter uma análise crítica de especialistas dos campos envolvidos e de outros que possam ter interesse no tema, especialmente dentro do contexto das bibliotecas.

A proposta de preservação tratada nesse tópico baseia-se em vários modelos, contudo alguns possuem mais relevância como o *Open Archival Information System (OAIS)* e o *the Cedars Project*, anteriormente citados no trabalho.

Basicamente o modelo da NLA foi criado como um sistema de saída de dados e não de entrada, ou seja a preocupação maior se dá com o que vai sair do sistema e não com as informações que serão introduzidas na criação do conjunto de dados e metadados e como se dará a associação desses arquivos. A idéia é determinar o que é relevante para que o sistema de preservação mantenha os dados para que cada ambiente os trabalhe e associe da maneira como melhor convier.

É importante ressaltar também que a iniciativa concentra-se apenas nos requisitos para a preservação e em nenhuma outra parcela do processo que possa ser igualmente importante para um processo mais amplo. Diferentes tipos de arquivos e sistemas de arquivamento distintos necessitam de metadados próprios e talvez não sejam contemplados dentro da iniciativa.

O conjunto de metadados baseia-se na necessidade de gerir e descrever coleções, objetos e sub-objetos (arquivos). Para cada nível existirá um conjunto específico de metadados relevantes. Espera-se que decisões pragmáticas determinem o nível de especificidade necessário

para cada caso. Segundo a NLA outro item destacado como principal em qualquer sistema de preservação é o histórico de preservação para tanto apresenta duas opções:

- Criar um único registro ao longo de todo o processo;
- Criar um novo registro a cada vez que o arquivo for modificado ou sofrer migração.

Acredita-se que o segundo processo seja mais interessante. A quantidade de informações geradas no processo de preservação dos objetos é muito grande. A criação de arquivos de metadados sobre a história a cada mudança vai organizar essa informação de maneira mais clara e conveniente. Os elementos recomendados são:

| Nome do Elemento | Definição |
|---|---|
| <i>URL persistente.</i> | Um nome que identifique o objeto de maneira única e persistente permitindo que links de diversos metadados sobre ele sejam direcionados. |
| <i>Data da Criação</i> | A data de criação do objeto. Esse dado cominado com outros metadados ajudará a fornecer evidência de autenticidade e proveniência para objetos. |
| <i>Tipo de Estrutura</i> | O tipo do objeto ou coleção é descrito usando as categorias a seguir: imagem, som, vídeo, texto base de dados, software, ou a cominação de dois ou mais como no caso de documentos WEB. Essa lista deve ser extensiva para acomodar novos objetos. |
| <i>Infra-estrutura técnica de objeto complexo</i> | Gerenciar a infra-estrutura técnica de objetos complexos bem como de seus componentes para contextualizar a informação (arqueologia digital). Inclui também aspectos técnicos de objetos ou coleção como formatos de arquivos, modo de compressão, dentre outros. |
| <i>Descrição de arquivo</i> | São recomendados campos de descrição para cada elemento segundo os tipos de arquivos o tipo de estrutura (Structural Type), anteriormente descrito. Apresenta-se abaixo um exemplo de metadados recomendados para arquivos de texto. |

Quadro.: Exemplo de elemento extraído da Preservation Metadata for Digital Collections Initiative

| Nome do elemento | Definição |
|---|--|
| 5.4.1 Formato de Texto e Versão | Definição: O Tipo do arquivo e sua versão Exemplo: MS Word 97 |
| 5.4.2 Compressão | Definição: O tipo e o nível de compressão Exemplo: .zip file |
| 5.4.3 Conjunto de Caracteres utilizados | Definição: O conjunto de caracteres usados do documento; Exemplos: ASCII; Unicode; EBCDIC |
| 5.4.4 Texto Associado | Definição: Nome do documento e tipo de definição aplicada para estruturar o texto. Exemplo: EAD |
| 5.4.5 Divisões estruturais de Texto | Definição: As divisões lógicas e estruturadas dos arquivos de texto Exemplos: TEI element DIVn used |

Quadro: Exemplo de definição Preservation Metadata for Digital Collections Initiative

| Nome | Definição |
|--|--|
| <i>Know System Requeriments (requisitos de sistema conhecidos)</i> | Os requisitos de sistema ou software necessários para acessar o objeto e utilizá-lo. Pode descrever a gama de sistemas necessários para que o objeto funcione e dar informações sobre as versões antigas e novas e de compatibilidade (se o objeto continua funcionando nas versões mais novas). |
| <i>Installation Requeriments (requisitos para instalação)</i> | Procedimentos específicos necessários para a instalação do objeto. É fundamental para objetos que tenham requisitos especiais de instalação |
| <i>Storage Information (informação para armazenamento)</i> | Capacidade de armazenamento para objetos e detalhes do sistema de armazenamento incluindo formatos físicos. |
| <i>Access Inhibitors (inibidor de Acesso)</i> | Qualquer método usado para inibir o acesso que impacte nos procedimentos de preservação como criptografia. |
| <i>Finding and Searching Aids, and Access Facilitators (ajuda para busca e facilitador de acesso)</i> | Qualquer método ou sistema usado para incrementar o acesso à informação dentro do objeto que precise ser mantido para sucessivas gerações. |
| <i>Preservation Action Permission (permissão para ação de preservação)</i> | Utilizado para determinar, sobre os aspectos legais ou institucionais, quem tem permissão para preservar os objetos digitais. |
| <i>Validation (validação)</i> | Informações sobre mecanismos de validação, seja do arquivo o do sistema arquivístico. |
| <i>Relationships (relacionamentos)</i> | Relacionamento entra o objeto e os demais necessários para o gerenciamento da preservação. |
| <i>Quirks (aparência)</i> | Qualquer característica que possa representar perda de funcionalidade ou mudança na aparência do arquivo |
| <i>Archiving Decision (work) Decisão para Arquivo</i> | Detalhes sobre o trabalho de arquivo deve ser incluído na data da decisão esse campo também pode incluir revisões. |
| <i>Decision Reason (work) Reação de Decisão</i> | Porque a decisão de arquivar ou não o trabalho foi tomada |
| <i>Institution Responsible for Archiving Decision (work) – (Responsabilidade institucional por decisão de arquivo)</i> | O nome da instituição responsável pela decisão de arquivar o trabalho. |
| <i>Archiving Decision (manifestation)- (Decisão por arquivo)</i> | A decisão se esse documentos deverá ser arquivado / conservado e a data em que foi deferida. |
| <i>Decision Reason (manifestation)- (Decisão por arquivo – Manifestação)</i> | Porque a decisão de arquivar ou não o objeto foi tomada. |
| <i>Institution Responsible for Archiving Decision (manifestation) Instituição Responsável pela decisão para arquivo (manifestação)</i> | O nome da instituição responsável pela decisão de que essa manifestação deve ser arquivada ou mantida. |
| <i>Intention Type (Tipo de intenção)</i> | A utilização de uma determinada manifestação de intenção em relação ao objeto. |

| | |
|---|--|
| <i>Institution with preservation responsibility (Instituição com a responsabilidade pela preservação)</i> | O nome da instituição que aceitou a responsabilidade pela preservação. Deverá incluir a data de início da aceitação e / ou os intervalos de datas de responsabilidade. |
| <i>Process (Processo)</i> | Todos os detalhes relevantes sobre o processo aplicado em um objeto digital ou arquivo, incluindo software, ações específicas, procedimentos, detalhes sobre equipamentos, pessoas ou instituições responsáveis. |
| <i>Record Creator (Criador do registro)</i> | O nome da instituição e das pessoas que contribuíram para com dados para o registro |
| <i>Other (outros)</i> | Qualquer outra informação relevante para a preservação da coleção, objeto ou arquivo. |

Quadro: Termos extraídos da Preservation Metadata for Digital Collections Initiative

5.6) METS-Metadata Encoding and Transmission Standard

A iniciativa *METS – Metadata Encoding Transmission Standards* é baseada no padrão OAIS e tem como objetivo criar um padrão de metadados em formato XML que possa gerenciar coleções de objetos digitais. Utiliza a formatação OAIS criando três estágios possíveis: Pacote de Informação de Submissão (*Submission information Package SIP*); pacote de informação de arquivo (*Archival Information Package AIP*); pacote de informação de disseminação (*Dissemination Information Package DIP*). A iniciativa é um padrão desenvolvido pelo *Network Development and MARC Standards Office* da *Library of Congress*. Possui Sete seções principais:

1. Cabeçalho METS: permite registrar dados sobre o objeto METS em si dentro de um documento METS. Esse metadados incluem informações sobre data de criação, de alteração de documento, responsáveis pela criação ou que tenham desempenhado algum papel dentro do documento.
2. Metadados Descritivos: essa seção pode apontar para metadados externos ao documento METS. Ou conter metadados descritivos embutidos ou ambos. Podem ser incluídas múltiplas seções de metadados descritivos.
3. Metadados Administrativos: Essa seção fornece informações sobre os arquivos que foram criados e armazenados. Informações sobre como o objeto foi criado ou armazenado, sua proveniência, direitos de propriedade intelectual dentre outros importantes para a administração do objeto. Podem ser armazenados interna ou externamente ao objeto.
4. Seção de Ficheiros: Contém a lista de todas as versões do objeto digital. Arquivos podem ser agrupados em grupos para permitir uma subdivisão por versão do objeto.

5. Mapa Estrutural: Coração do documento METS. Esboça uma estrutura hierárquica para o objeto da biblioteca digital e liga elementos dessa estrutura a outros objetos (fichas) que contenham metadados sobre o elemento.
6. Ligações Estruturais: Permite registrar a existência de hiperlinks na estrutura do documento METS e no mapa estrutural. Especialmente importante no caso de *sites*.
7. Comportamento: Associa comportamento executáveis como conteúdo do objeto METS. Comportamentos remetem a uma determinada ação ligada a um documento como por exemplo uma ligação com um conteúdo on-line.

Em resumo a idéia do METS é oferecer um mecanismo flexível para codificar metadados descritivos, administrativos e estruturais para objetos de bibliotecas digitais. Além disso, procura descrever a ligação entre objetos e oferecer uma norma útil para a troca de objetos entre repositórios.

5.7) NOBRADE – Norma Brasileira de Descrição Arquivística

Por fim e igualmente importante buscou-se estudar a NOBRADE – Norma Brasileira de Descrição Arquivística haja vista que esse trabalho lida especialmente com metadados para preservação no âmbito da arquivística. A referida norma baseia-se no padrão internacional *ISAD - International Standard Archival Description*, traduzida para o português pelo Conselho Nacional de Arquivos sob a denominação: ISAd(G) – Norma Geral Internacional de descrição Arquivística.

A iniciativa herda da *ISAD* uma estrutura hierárquica de níveis que remete aos princípios de proveniência e organicidade, pilares da teoria arquivística, citados alhures nesse trabalho. Considera-se a existência de seis níveis principais de descrição: acervo da entidade custodiadora (nível 0), fundo ou coleção (nível 1), seção (nível 2), série (nível 3), dossiê ou processo (nível 4) e item documental (nível 5). São admitidos como níveis intermediários o acervo da subunidade custodiadora (nível 0,5), a subseção (nível 2,5) e a subsérie (nível 3,5).

A NOBRADE prevê a existência de oito áreas, para cada nível descritivo, que podem ser tratadas como classes de informação que no total compreendem 28 elementos de descrição. As áreas, conforme denominado pelo Conarq constituem-se em:

- Área de Descrição: onde se registra informações essenciais sobre a unidade de descrição;
- Área de Contextualização: registram-se informações sobre a proveniência e custódia da unidade de descrição;

- Área de Conteúdo e Estrutura: onde se registra informações sobre o assunto e a organização da unidade de descrição;
- Área de condições de acesso: onde se registra informações sobre o acesso à unidade de descrição;
- Área de Fontes Relacionadas: Registram-se informações sobre outras fontes que tem importante relação com a unidade de descrição;
- Área de Notas: Informações sobre o estado de conservação e / ou qualquer outra informação sobre a unidade de descrição que não tenha lugar nas áreas anteriores;
- Área de Controle da Descrição: Onde se registra Informação sobre como quando e por quem a descrição foi elaborada;
- Áreas de Ponto de acesso e descrição de assuntos: onde se registra informações sobre termos selecionados para localização e recuperação da unidade de descrição.

Dos vinte e oito elementos de descrição disponíveis, sete são obrigatórios:

- Código de referência
- Título
- Data(s)
- Nível de Descrição
- Dimensão e Suporte
- Nome(s) dos(s) Produtor(es)
- Condições de Acesso

Descreve-se abaixo o conjunto simplificados de termos relacionados a NOBRADE:

Área de Identificação

- Código de Referência – Identificar a unidade de descrição;
- Título – Identificar nominalmente a unidade de descrição;
- Data – Informar a(s) data(s) da unidade de descrição;
- Data Tópica;
- Data Crônica;
- Data de Produção;
- Data de Acumulação;
- Data-Assunto;
- Nível de descrição – Identificar o nível da unidade de descrição em relação às demais;
- Dimensão e Suporte – Identifica as dimensões físicas ou lógicas da unidade de descrição, relacionando esse dado ao respectivo suporte;

- Área de contextualização;
- Nomes dos Produtores – Identificar os produtores da unidade de descrição;
- Identificação do produtor;
- Nível de descrição;
- Indicação do Nome do Produtor;
- História administrativa / biografia – Oferecer informações referenciais sistematizadas da trajetória dos produtores da sua criação ou nascimento até sua extinção ou falecimento.

História Arquivística - Objetivo: Oferecer informações referenciais sistematizadas sobre a história da produção e armazenamento do documento

- Procedência – Identificar a origem imediata de aquisição ou transferência da unidade de descrição;
- Área de Conteúdo e Estrutura;
- Âmbito e Conteúdo – Oferecer, aos usuários, informações relevantes ou complementares ao título;
- Avaliação, eliminação e temporalidade – Fornecer informações sobre qualquer ação relativa à avaliação, seleção e eliminação;
- Incorporações – Informar aos usuários sobre acréscimos previstos à unidade de descrição;
- Indicação de previsão de incorporação.;

Sistema de Arranjo – Fornecer informação sobre a estrutura interna e / ou sistema de arranjo da unidade de descrição.

- Área de condições e acesso de uso;
- Condições de Acesso – Fornecer informação sobre as condições de acesso a unidade de descrição e, existindo restrições, em que estatutos legais ou outros regulamentos se baseiam;
- Condições de reprodução – identificar qualquer restrição quanto a reprodução da unidade de descrição;
- Idioma – identificar os idiomas, escritas e sistemas de símbolos utilizados na unidade de descrição;
- Características físicas e requisitos técnicos – Informe características físicas requisitos técnicos importantes que afetem o uso da unidade de descrição;
- Características Físicas;
- Equipamentos e meios necessários para uso da unidade de descrição.

- Instrumentos de pesquisa – Identificar os instrumentos de pesquisa relativos a unidade de descrição;
- Áreas de Fontes relacionadas;
- Existência e localização dos originais – Indicar a existência e localização, ou inexistência, dos originais de uma unidade de descrição constituída por cópias;
- Existência e localização de cópias – Indicar a existência e localização de cópias da unidade de descrição;
- Unidades de descrição relacionadas – Identificar a existência de unidades de descrição relacionadas;
- Nota sobre publicação – identificar Publicações sobre a unidade de descrição ou elaboradas com base no seu uso, estudo e análise bem como, as que referenciem, transcrevam ou reproduzam;
- Área de Notas
 - Notas sobre conservação
- Notas gerais
- Área de Controle da descrição
- Nota do arquivista
- Fontes utilizadas
- Códigos de referência dos documentos utilizados;
- Nome e responsabilidades das pessoas envolvidas no trabalho
- Regras ou Convenções – Identificar as normas e convenções em que a descrição é baseada.
- Data das descrições
- Data em que a descrição foi preparada e / ou revisada, de maneira cumulativa ou não, preservando aquela inicial e da última revisão.
- Área de Pontos de acesso e indexação de assuntos
- Pontos de acesso e indexação de assuntos – Registrar os procedimentos para recuperação do conteúdo de determinados elementos de descrição, por meio da geração e elaboração de índices baseados em entradas autorizadas e no controle do vocabulário adotado.

A norma brasileira de descrição arquivística vem sendo usada em inúmeros projetos como fonte de descrição de recursos informacionais importante dentro da realidade brasileira. Sua inclusão dentro das iniciativas pesquisadas é importante na medida em que

representa uma importante fonte arquivística e representa, em grande medida, a realidade arquivística brasileira dentro do contexto do Conselho Nacional de Arquivos.

Considerando-se os seis níveis principais, multiplicados a oito áreas e vinte e oito elementos obtém alguma exaustividade. O intuito desse trabalho é retirar as entidades para, em conjunto com as demais iniciativas, compor um conjunto de termos para o estado terminológico planejado.

5.8) Implicações para a pesquisa

A presente sessão apresentou algumas iniciativas reconhecidas internacionalmente no âmbito da preservação digital, além da Norma Brasileira de Descrição Arquivística que ajuda a integrar os conceitos da arquivística.

Cada iniciativa possui características distintas, pois trata de um assunto diferente ou possui um propósito diferente. Isso foi muito importante para o trabalho devido aos diversos pontos de vista.. Cada iniciativa foi feita com base em autoridades reconhecidas no mundo da informação o que fornece credibilidade ao trabalho na medida em que via de regra buscou-se inserir na tabela de estado terminológico todos os termos tratados como obrigatórios nas iniciativas além daqueles que estivessem diretamente ligados a arquivologia. Optou-se ainda por incluir os termos que constam em mais de uma iniciativa por entendimento de que se torna importante em dois ou mais contextos.

Admite-se uma pequena subjetividade para suprimir ou inserir alguns raros termos que não se enquadrem nesses dois princípios considerando-se boas práticas adotadas na arquivologia oriundas de experiências práticas de mercado, contudo ressalta-se que as exceções não chegam a dois por cento dos termos. Admite-se ainda que os termos não devem ser tomados como referência absoluta para construção de trabalhos que, até pelo caráter singular da arquivística podem ser utilizados ou não dentro do contexto informacional de cada instituição.

Em resumo pode-se dizer que o estudo das iniciativas foi a base para o levantamento dos termos e construção da tabela de estado terminológico permitindo assim a tomada de decisões objetivas em relação a escolha dos mesmos.

6. RESULTADOS

A presente seção aborda a definição do estágio terminológico preliminar, conforme mencionado na seção 1, que se compõe de termos representativos de metadados para preservação digital. Buscou-se reunir fundamentação teórica sobre arquivologia, preservação de registros digitais assim como a realização de um estudo que envolveu seis iniciativas mundiais para preservação de registros digitais, além da Norma Brasileira de Descrição arquivística.

A partir desses estudos reuniu-se um conjunto de procedimentos, requisitos e especialmente um conjunto de campos, aqui tratados enquanto metadados, fundamentais para garantir a preservação dos registros digitais ao longo dos anos ou até mesmo décadas. Foram levantadas ainda informações acerca dos modelos construídos a partir de princípios ontológicos o que levou a definir um modelo adequado aos propósitos deste trabalho.

Todas as iniciativas levantadas advêm de instituições mundialmente reconhecidas e que já estão implantando projetos de preservação de registros digitais através de longos processos de discussão que envolvem uma série de especialistas numa equipe de pesquisa interdisciplinar. Entende-se que o uso de padrões e de iniciativas mundialmente reconhecidas é fundamental não somente para garantir a interoperabilidade do trabalho e a sustentabilidade do mesmo enquanto processo científico dentro da proposta de pesquisa de natureza aplicada; de abordagem qualitativa e de caráter experimental.

A que se levar em consideração que cada iniciativa possui um propósito específico e nem sempre os termos definidos se enquadram dentro de um caráter mais geral que visa apenas a preservação dos registros digitais como é o caso dessa ontologia. Buscou-se reduzir a subjetividade na escolha dos termos. Para tanto se considerou os termos essenciais ou obrigatórios dentro das iniciativas. Foram incluídos ainda aqueles que se repetem ainda que tenham uma terminologia diferente, mas que remetam ao mesmo sentido. Em alguns casos considerou-se a inclusão de termos que não são oriundos das iniciativas estudadas, mas que se aplicam ao trabalho arquivístico e são importantes para contextualizar o conjunto de metadados dentro da terminologia em desenvolvimento. Nesse caso no local, onde estaria o nome da iniciativa de origem, consta um traço “-“.

Leva-se em conta ainda a estrutura em níveis utilizada pela Norma Brasileira de descrição arquivística previamente descrita nesse trabalho. Nesse sentido deve-se observar, especialmente dentro do âmbito da arquivística, que cada fundo, classe, sub-classe ou objeto em si deve ser descrito. O maior ou menor nível de detalhamento deve ser previsto dentro da política de informações da instituição.

Utiliza-se para efeitos desse trabalho uma estrutura hierárquica que se relaciona e define termos e procedimentos. Criam-se também tabelas auxiliares com definições, equivalências e adaptações de termos para a realidade proposta.

A tabela com o estado terminológico dos metadados inclui a origem, nome em sua iniciativa original, o termo utilizado no contexto desse trabalho, sua definição. Em virtude da ampla extensão dos resultados, tal tabela com os resultado finais do estágio terminológico preliminar foi transferida para o Anexo 2. No restante da presente seção, explica-se como a terminologia preliminar é organizada de acordo com princípios ontológicos do CIDOC-CRM.

Para organização da terminologia, considera-se um subconjunto da CIDOC-CRM, denominado CRM-*Compatible Form* (CROFTS et al, 2003). Em seguida apresenta-se a estrutura compatível original, como as entidades de alto nível em inglês (CROFTS et al, 2003, p.11), para então, adicionar os termos levantados na Seção 5 e constantes no Anexo 2. (Fig. 21). A CIDOC é apresentada em inglês, a linguagem natural original em que foi desenvolvida.

| | |
|------------|--|
| E1 | CRM Entity |
| E2 | - Temporal Entity |
| E4 | - - Period |
| E5 | - - - Event |
| E7 | - - - - Activity |
| E11 | - - - - - Modification |
| E12 | - - - - - - Production |
| E13 | - - - - - - Attribute Assignment |
| E65 | - - - - - - Creation |
| E63 | - - - - - - Beginning of Existence |
| <i>E12</i> | - - - - - - <i>Production</i> |
| E65 | - - - - - - Creation |
| E64 | - - - - - - End of Existence |
| E77 | - Persistent Item |
| E70 | - - Thing |
| E72 | - - - Legal Object |
| E18 | - - - - Physical Thing |
| E24 | - - - - - Physical Man-Made Thing |
| E90 | - - - - - Symbolic Object |
| E71 | - - - - - Man-Made Thing |
| <i>E24</i> | - - - - - <i>Physical Man-Made Thing</i> |
| E28 | - - - - - Conceptual Object |
| E89 | - - - - - - Propositional Object |
| E30 | - - - - - - - Right |
| E73 | - - - - - - - Information Object |
| <i>E90</i> | - - - - - - - <i>Symbolic Object</i> |
| E41 | - - - - - - - Appellation |
| <i>E73</i> | - - - - - - - <i>Information Object</i> |
| E55 | - - - - - - - Type |
| E39 | - - Actor |
| E74 | - - - Group |
| E52 | - Time-Span |
| E53 | - Place |
| E54 | - Dimension |

E59 Primitive Value
 E61 - Time Primitive
 E62 - String

| Property id | Property Name | Entity – Domain | Entity – Range |
|-------------|---|-----------------------------|------------------------------------|
| P1 | is identified by (identifies) | E1 CRM Entity | E41 Appellation |
| P2 | has type (is type of) | E1 CRM Entity | E55 Type |
| P3 | has note | E1 CRM Entity | E62 String |
| P4 | has time-span (is time-span of) | E2 Temporal Entity | E52 Time-Span |
| P7 | took place at (witnessed) | E4 Period | E53 Place |
| P10 | falls within (contains) | E4 Period | E4 Period |
| P12 | occurred in the presence of (was present at) | E5 Event | E77 Persistent Item |
| P11 | - had participant (participated in) | E5 Event | E39 Actor |
| P14 | - - carried out by (performed) | E7 Activity | E39 Actor |
| P16 | - used specific object (was used for) | E7 Activity | E70 Thing |
| P31 | - has modified (was modified by) | E11 Modification | E24 Physical Man-Made Thing |
| P108 | - - has produced (was produced by) | E12 Production | E24 Physical Man-Made Thing |
| P92 | - brought into existence (was brought into existence by) | E63 Beginning of Existence | E77 Persistent Item |
| <i>P108</i> | - - <i>has produced (was produced by)</i> | <i>E12 Production</i> | <i>E24 Physical Man-Made Thing</i> |
| P94 | - - has created (was created by) | E65 Creation | E28 Conceptual Object |
| P93 | - took out of existence (was taken out of existence by) | E64 End of Existence | E77 Persistent Item |
| P15 | was influenced by (influenced) | E7 Activity | E1 CRM Entity |
| <i>P16</i> | - <i>used specific object (was used for)</i> | <i>E7 Activity</i> | <i>E70 Thing</i> |
| P20 | had specific purpose (was purpose of) | E7 Activity | E7 Activity |
| P43 | has dimension (is dimension of) | E70 Thing | E54 Dimension |
| P46 | is composed of (forms part of) | E18 Physical Thing | E18 Physical Thing |
| P59 | has section (is located on or within) | E18 Physical Thing | E53 Place |
| P67 | refers to (is referred to by) | E89 Propositional Object | E1 CRM Entity |
| P75 | possesses (is possessed by) | E39 Actor | E30 Right |
| P81 | ongoing throughout | E52 Time-Span | E61 Time Primitive |
| P82 | at some time within | E52 Time-Span | E61 Time Primitive |
| P89 | falls within (contains) | E53 Place | E53 Place |
| P104 | is subject to (applies to) | E72 Legal Object | E30 Right |
| P106 | is composed of (forms part of) | E90 Symbolic Object | E90 Symbolic Object |
| P107 | has current or former member (is current or former member of) | E74 Group | E39 Actor |
| P127 | has broader term (has narrower term) | E55 Type | E55 Type |
| P128 | carries (is carried by) | E24 Physical Man-Made Thing | E73 Information Object |
| P130 | shows features of (features are also found on) | E70 Thing | E70 Thing |
| P140 | assigned attribute to (was attributed by) | E13 Attribute Assignment | E1 CRM Entity |
| P141 | assigned (was assigned by) | E13 Attribute Assigment | E1 CRM Entity |
| P148 | has component (is component of) | E89 Propositional Object | E89 Propositional Object |

Figura 20: Forma compatível do CRM

A inclusão dos temas na estrutura do CIDOC-CRM permitiu obter o alinhamento necessário para continuidade da especificação de metadados digitais, de forma a interoperar com outras iniciativas. Para efeito da metodologia seguida, resultados são apresentados na estrutura hierárquica da CIDOC. Os termos em caixa escura são os originais e os termos em caixa clara os adicionados, provenientes da extração de termos realizada na seção 5 e apresentada no anexo 2.

| | |
|-----|--|
| E1 | CRM Entity |
| E2 | - Temporal Entity |
| E4 | - - Period |
| E5 | - - - Event |
| E7 | - - - - Activity |
| E11 | - - - - - Modification |
| E12 | - - - - - Production |
| | - Processo de digitalização |
| | - Metodo de compressão |
| | - Método de criptografia |
| | - |
| E13 | - - - - - Attribute Assignment |
| | - Identificação |
| | - Identificação do objeto individual |
| | - Tipo de objeto |
| | - Código de representação |
| | - Código identificador do objeto |
| | - Código identificador do domínio |
| | - Aspecto visual do objeto |
| | - Aparência |
| | - Qualidade da fontes de dados |
| | - Data de uso |
| E65 | - - - - - Creation |
| | - Ação |
| | - Relacionamento estrutural |
| | Consequencia do uso para efeitos probatórios |
| E63 | - - - - - Beginning of Existence |
| E12 | - - - - - <i>Production</i> |
| E65 | - - - - - Creation |
| E64 | - - - - - End of Existence |
| E77 | - Persistent Item |
| E70 | - - Thing |
| E72 | - - - Legal Object |
| E18 | - - - - Physical Thing |
| E24 | - - - - - Physical Man-Made Thing |
| | - Pré requisitos técnicos |
| | - Documentação |
| | - Padrão de documentação do objeto |
| | - Normas para ligação entre hipertexto |
| | - Padrão de intercâmbio entre objetos |
| | - Instrumento de captura dos dados |
| E90 | - - - - - Symbolic Object |
| E71 | - - - Man-Made Thing |
| E24 | - - - - <i>Physical Man-Made Thing</i> |
| E28 | - - - - Conceptual Object |
| E89 | - - - - - Propositional Object |
| E30 | - - - - - Right |
| | - Copyright |
| | - Informações sobre direitos autorais |
| | - Legislação relacionada |

| | | | |
|------|---|-----------------|---------------------------------|
| | | - | Condições de acesso |
| | | - - | Condições de técnicas de acesso |
| E54 | - | Dimension | |
| E 55 | - | Language | |
| | | - | Idioma |
| E59 | | Primitive Value | |
| E61 | - | Time Primitive | |
| E62 | - | String | |

Figura 21: inserção termos na estrutura do CIDOC

As relações semânticas entre os termos utilizados na terminologia preliminar não são aqui detalhadas, uma vez que suas propriedades são herdadas diretamente das propriedades da CIDOC-CRM. Detalhar tais relações significaria repetir as distinções rigorosas e disponíveis na documentação da CIDOC.

7. CONSIDERAÇÕES FINAIS

No capítulo de introdução foram apresentadas questões de pesquisa. Nesse capítulo busca-se responder as questões mediante as pesquisas efetuadas no decorrer desse trabalho:

- i) a utilização de metadados é efetiva como forma de garantir a preservação dos documentos digitais a longo prazo?
- ii) quais as relações entre a arquivística tradicional e a nova realidade que são benéficas para a preservação de documentos digitais a longo prazo?
- iii) a inexistência de conjuntos de metadados inviabiliza a preservação?
- iv) o uso de princípios ontológicos para organização de termos representativos de metadados pode auxiliar na construção de instrumentos que auxiliem na preservação de documentos digitais?

A resposta à **primeira questão** fica clara em diversos pontos da pesquisa. Não somente os metadados são efetivos, na medida em que apresentam resultados reais no processo de preservação, como também são parte desse processo. Conforme descrito na seção 3.2 somente o uso de metadados pode garantir o princípio da proveniência no meio digital. Na seção 3.3 são descritas as técnicas e estratégias para preservação de documentos digitais. São diversas técnicas e em todos os casos os metadados são fundamentais seja no apoio direto de armazenamento de dados ou no apoio indireto para recuperação do contexto ou dos próprios dados. Destaca-se ainda que qualquer modificação nos objetos digitais causa alteração dos metadados o que os torna parte fundamental de todo o processo.

A importância vital dos metadados é evidenciada pela norma ISO 19005-1 citada nesse trabalho que determina o uso do formato PDF/A e destaca a possibilidade de criação de um suporte interno a metadados no caso de imagens comprimidas. Tal formato, por natureza, faz uma separação do conteúdo bruto do arquivo e de seus metadados o que pode indicar um caminho futuro na preservação já que parte do problema é justamente a interpretação dos arquivos por softwares diferentes que se dá especialmente pela mudança na interpretação dos metadados.

Os metadados estão novamente em destaque na medida em que todo documento digital os gera de maneira automática e sob diversos aspectos. Imaginemos um texto editado em um software apropriado. Além dos caracteres de texto ele tem que conter as propriedades do editor, do sistema operacional, do sistema de armazenamento, dentre outros. Para tanto a criação

ou o uso de padrões pré-existentes como *Dublin Core*, dentre tantos outros pode ser um caminho facilitador no intuito de diminuir os problemas de armazenamento e preservação já que sempre serão utilizados os mesmos critérios. Além dos critérios todas as iniciativas pesquisadas apontam para a necessidade da criação e manutenção de metadados para efeitos de preservação.

Pode-se dizer que durante toda a história tentou-se preservar os suportes de informação na medida em que os dados ou informações ali presentes estavam intrinsecamente ligados a os mesmos. Percebe-se que com o advento do suporte digital o foco mudou. Conforme tratado no decorrer do trabalho o foco não é mais o objeto, mas a garantia de acesso dentro de um universo complexo de objetos binários distintos, definidos e interligados por uma complexa teia de metadados.

Por fim, concluindo a primeira pergunta de pesquisa pode-se dizer que a utilização de metadados para efeitos de preservação digital é tratada como extremamente importante pelas iniciativas estudadas nesse trabalho, que tem como autores respeitadas instituições e um trabalho interdisciplinar. São desenvolvidas por pesquisadores e profissionais conceituados e nesse sentido apresenta relevância consensual.

Em relação à **segunda questão**, que trata sobre a existência de relação entre a arquivística tradicional e a preservação digital, pode-se responder que existe relação forte entre os arquivos digitais e tradicionais. Em verdade, os princípios são os mesmos previamente estabelecidos apenas a maneira de realizar os procedimentos serão alterados principalmente pela inclusão de metadados.

Em princípio pode parecer que o advento da informática traga uma solução arquivística que dependa exclusivamente de um software ou das ferramentas da Ciência da Informação para garantir os princípios arquivísticos. No âmbito do senso comum isso tem se tornado algo bastante difundido. Percebe-se, por exemplo, que as bibliotecas virtuais e tradicionais, perdem grande espaço para a internet e, diversos estudos demonstram que alunos do ensino primário já desconsideram o uso de fontes de informação que não estejam plugadas na grande rede. A arquivologia não fica fora dessa visão. Teóricos há que defendem que, num futuro próximo, a informática suprirá todas as funções da arquivologia, teoria essa que não é confirmada por esse trabalho, que aponta um caminho contrário na medida em que se confirma a necessidade da manutenção dos requisitos arquivísticos independente do suporte.

O que acontece é que em muitos casos, conforme citado no decorrer do trabalho, e ratificado por Malheiro (2001), o profissional de informática vem substituindo o profissional da informação de maneira deficiente. Isso se dá em especial no âmbito da gestão de documentos digitais. Isso se configura em um engano já que a gestão dos documentos eletrônicos, conforme

se percebe no decorrer da pesquisa, está essencialmente ligada à recuperação da informação e seus desdobramentos como a preservação digital, cerne desse trabalho. Obviamente existe uma tendência forte a transdisciplinaridade, mas a colaboração deve dar-se de maneira conjunta respeitando os limites de cada área do conhecimento.

Em verdade pode-se dizer que dentro do contexto digital os princípios arquivísticos, utilizados em meio físico, não somente são válidos como se tornam ainda mais importantes. Justifica-se tal afirmativa recorrendo a princípios arquivísticos: o primeiro enfatiza os documentos de valor primário, ou seja, diretamente ligados a administração ou ao fim que os criou; numa segunda considera-se, me destaque, os documentos com valor histórico ou secundário, e numa terceira trata-se de uma arquivologia de caráter global que trata de todos os documentos.

Sendo a terceira visão a mais utilizada e mais acertada dentro da perspectiva desse trabalho o meio digital é uma importante ferramenta no que diz respeito ao conceito de gerenciamento de arquivos que vai tratar de todo o trâmite documental desde a criação até o descarte ou sua preservação. Seria impossível realizar todo esse trabalho sem valer-se dos conceitos aplicados aos documentos tradicionais na medida em que a perda do contexto documental traria prejuízos irreversíveis e até mesmo descaracterizaria os mesmos.

O ambiente digital traz grandes facilidades para o tratamento da informação. Podem-se utilizar diversos recursos que seriam impossíveis no meio físico, contudo para gerenciar documentos arquivísticos no meio digital deve-se levar em conta um nível organizacional ainda maior, pois as aparentes facilidades geram confusões ainda maiores, pela constante multiplicação dos documentos, dos acessos simultâneos e a falta de barreiras físicas. Nesse ambiente perder a proveniência de um documento, por exemplo, pode descaracterizá-lo e torná-lo não fidedigno. Nesse sentido todos os princípios arquivístico utilizados em meio físico devem ser observados em meio digital, contudo, isso se dá de maneira lógica pelo controle de softwares especializados e essencialmente por metadados.

Sem tais conceitos os documentos arquivísticos, independente do suporte em que estão armazenados, perderiam seu valor intrínseco que, segundo o Arquivo Nacional (2005) deve ser de prova dentro dos contextos legal, fiscal, probatório ou administrativo além de seu contexto que é garantido pelo princípio da proveniência. Documentos de arquivo possuem características únicas e por isso possuem princípios que não podem ser descartados independente do suporte utilizado.

O próprio conceito de documento de arquivo fica ameaçado se a tecnologia não estiver aliada à arquivística. O conceito de documento arquivístico remete a autenticidade e

fidedignidade. Embora os recursos de certificação possam garantir a inviolabilidade do documento somente através de um conjunto de relações lógicas previstas dentro da arquivística pode-se garantir que o documento é o que afirma ser. Para cada ato, um documento essa é a lei.

Outro grave problema na atualidade é o excesso de informação que sobrecarrega os servidores das empresas. Muito embora exista uma tendência de que o espaço digital torna-se cada vez mais abundante e mais barato por um movimento de avanço do estado da arte da tecnologia a busca pela informação dentro de um ambiente poluído torna-se cada vez mais complexa. Um grande exemplo disso são os sites de busca que possuem alta taxa de revocação, mas baixíssima precisão. O conceito da teoria das três idades pode desafogar os sistemas transferindo, segundo uma Tabela de Temporalidade Documental, os documentos que não sirvam diretamente para a administração para outros espaços tornando o sistema mais eficiente.

Pode-se defender ainda o uso de Plano de Classificação como forma de garantir que os documentos estejam estruturados segundo a maneira como foram gerados. É bastante comum que computadores pessoais e empresariais tenham uma infinidade de documentos soltos e sem contexto que acabam se perdendo ou gerando informações errôneas ou desatualizadas. No arquivo a classificação é orgânica e não filosófica como em outros casos. Esse é um erro comum quando documentos de arquivo são classificados levando-se em conta somente a tecnologia.

A grande mudança é que dentro do meio tradicional tais princípios são garantidos muito mais pela disposição física dos documentos e no meio digital são os metadados que vão fazer esse trabalho. Logo se conclui que os princípios continuam válidos mudando apenas a maneira de aplicação. A manipulação dos documentos arquivísticos passíveis de preservação deve ocorrer sob a ótica dessa área do conhecimento valendo-se para tanto dos mesmos princípios utilizados no meio físico.

Mudam-se os suportes, mas continuamos com os mesmos princípios. A idéia de que se pode dispensar as teorias é oriunda de uma “aura mágica” que traz um caráter mágico para a tecnologia que parece resolver todos os problemas em si mesma, o que não é verdade. Os instrumentos, como a Tabela de Temporalidade Documental, possuem, dentre outras funções, o objetivo de minimizar a subjetividade dos processos determinando como cada situação deve ser tratada segundo política informacional previamente determinada por profissional qualificado. A importância dos princípios arquivísticos fica mais evidenciada quando se pensa no processo a longo prazo. Por mais que um documento, físico ou digital, seja preservado, como comprovar sua autenticidade, fidedignidade ou mesmo entender o seu contexto após décadas sem os princípios arquivísticos?

Respeitados, os instrumentos de arquivo, a informática pode ser poderosa aliada e grande ferramenta. Controle de fluxo de documentos, de suas versões, garantia de unicidade dentre tantos outros podem ser feitas de maneira mais eficiente dentro do ambiente digital. Destaque também para os processos de descrição que ganham possibilidades nunca imaginadas no ambiente físico. Os velhos índices e catálogos são substituídos por *softwares* que realizam combinações poderosas a fim de aumentar a precisão das buscas. Essa nova realidade traz a necessidade de criação de campos que excedem o ambiente tradicional, mas esses campos são necessários para que a própria tecnologia se faça presente pelas diversas razões técnicas amplamente discutidas durante o trabalho.

A **terceira pergunta** trata sobre a inexistência de conjuntos de metadados como fator de inviabilização da preservação. A resposta a essa questão foi evidenciada em diversos pontos do trabalho e abordada, de maneira indireta em sua conclusão, já que se enfatiza ali os metadados como fator preponderante em diversos momentos e processos de arquivos, especialmente no âmbito digital. Busca-se argumentar uma resposta positiva a esse questionamento no decorrer dessa resposta. Antes de mais nada é importante recorrer ao contexto de metadados que são dados que geram registros estruturados sobre outros dados e nesse sentido pode ser entendido como tudo o que está direta, indireta, intrínseca ou extrinsecamente ligado ao objeto a ser preservado.

Os processos que envolvem a preservação digital mostram-se muito mais complexos do que aparentam ser. Diversas técnicas devem ser combinadas e muitos fatores apresentam-se colocando-nos frente a uma realidade que além de extremamente complexa é dialética, pois muitas vezes o estado da arte da tecnologia parece ter uma lógica própria que acompanha quase sempre o mercado e não o bom senso.

Os metadados estão intrínsecos a esse processo. Consideramos que a base da informática moderna, da geração de computadores pessoais, é o Unicodeⁱ, Um texto recém digitado possui, por ele mesmo apenas caracteres de texto puro. Ou seja pode ser perfeitamente representado dentro do código básico, previamente citado. Contudo para que seja gerado ele depende de um sistema operacional, um usuário dentro de um contexto, e na maioria dos casos de um editor de textos que pode ser encontrado em diversos formatos. Todos essas fatores adicionais são os metadados. Logo os mesmos não podem ser separados, pois estão intimamente ligados.

Além dessa classe de metadados que são intrínsecos ao documento existem aqueles que serão criados para que o contexto do documentos em si seja preservado. São de ordem técnica, arquivística ou de criação de algum tipo de contexto, mas são metadados. Sem eles o

documentos torna-se algo perdido dentro de um espaço de informação e somente poderá ter utilidade para quem o criou ou para um grupo restrito de pessoas dentro de um espaço pequeno de tempo já que seu contexto será perdido brevemente.

Sendo objeto de estudo dessa pesquisa e cerne da pergunta em questão buscamos aprofundamento na questão daqueles metadados inerentes ao processo de preservação dos registros digitais. Durante o trabalho os metadados foram apresentados sob diversos ângulos e em todos os casos eles se mostraram fundamentais para o processo de preservação documental.

Antes de mais nada, a preservação digital deve estar inserida dentro de uma política de informação. Essa política será registrada através de um processo que envolve uma equipe interdisciplinar e a sistematização de regras que visam tornar cada ato objetivo e previamente definido. Isso será provavelmente o primeiro conjunto de metadados fundamentais para o processo de organização. Serão requisitos para registro das decisões, políticas e contexto de tudo o que for guardado.

O contexto deve ser preservado juntamente com os documentos para que no futuro se saiba como aqueles documentos formam uma teia de informações que caracterizam o documento de arquivo. Isso se dá fundamentalmente através de metadados. A falta de preocupação com esses metadados está diretamente ligada a perda de memória das instituições e sociedades que pode levar a curto prazo a uma lacuna pela falta de registros. Isso se dá tanto no âmbito técnico onde a falta de dados sobre os sistemas e estado da arte dos objetos leva a insuficiência de dados na sua recuperação tanto pelo hardware e pelo software quanto para o contexto geral do mesmo. Em muitos casos existe um processo de migração, mas feito de maneira mecânica e sem preocupação com os metadados o que leva os usuários do futuro a fazerem verdadeiras escavações digitais para decifrar os códigos binários.

O fato atenua-se se considerarmos que os arqueólogos trabalham com matérias que muitas vezes tem milhões de anos, no caso dos registros digitais estamos falando de no máximo algumas décadas. Novamente os metadados aparecem de maneira crucial dentro do contexto. Somente com a descrição pormenorizada dos objetos pode-se trabalhar os conceitos de preservação a longo prazo. Pode-se dizer que a parte lógica da preservação é totalmente dependente dos metadados a fim de trazer contexto adequado dentro do ambiente de recuperação e até mesmo para defini-lo.

São os metadados que garantem que um documento não será modificado, que sua transmissão é segura, que seu autor é quem diz ser. A própria legislação se apóia em metadados e sistemas para prover autenticidade a um documento. Exemplo do ICP – Instituto de Chaves Públicas, as autoridades certificadoras em resumo apóiam-se em sistema de metadados para

garantir que o autor seja identificado e que o processo que gerou o documento seja identificado, condições fundamentais para o efeito de prova.

Analisando os diversos contextos que definem o ambiente digital vemos que todos eles, muito além dos documentos em si, apóia em metadados. A própria existência do documento digital é ameaçada pela ausência de metadados, pois em teoria, tal objeto é virtual, ou seja, não existe até que possa ser decodificado por um conjunto de padrões pré estabelecidos para que possa ser apresentado num formato inteligível a percepção humana. Antes disso ele é um conjunto denominado *Bit Stream*ⁱⁱ. A cada nova etapa o documento recebe novos metadados e isso deve ser preservado afim de que o mesmo possa ser decodificado após sua guarda nos sistemas previamente programados para isso.

Pode-se dizer ainda que o uso de metadados não é uma grande novidade. Em verdade mesmo nos arquivos físicos isso era fundamental. O plano de classificação tanto no âmbito pré como pós coordenado representam metadados fundamentais para a contextualização dos documentos. O que a era digital trouxe foi incluir novos metadados em sua maioria fundamentais para garantir o entendimento do novo suporte, sem contudo dispensar os antigos. Nesse sentido traça-se um paralelo entre os dois suportes onde o uso de metadados é fundamental para os processos de preservação.

Um objeto digital possui uma grande teia de recursos necessários a sua visualização. Um similar preservado em sua forma física mantém todos os seus aspectos, desconsiderando-se aqui as particularidades do contexto arquivístico. No caso do documento digital os documentos podem ser facilmente decompostos caso não se tenha um cuidado de registrar, através de metadados, a relação complexa que existe, por exemplo, numa página da internet. Fala-se de uma mudança de ótica que sai do documento para o acesso do mesmo a longo prazo. Não interessa mais o documento, mas sua acessibilidade.

Pode-se afirmar que metadados são de suma importância dentro do processo de preservação digital, independente das estratégias e políticas definidas. Eles tem sido apontados como ferramenta fundamental para a preservação digital em todas as iniciativas estudadas no decorrer do trabalho e em alguns casos representam o objeto de pesquisa das mesmas seja pela definição de padrões ou pela definição dos critérios que devem basear as escolhas desses.

A **quarta e última pergunta** proposta diz respeito ao uso de princípios ontológicos para organização de termos representativos de metadados, como auxílio na construção de instrumentos que auxiliem na preservação de documentos digitais. Busca-se responder positivamente a questão segundo as questões levantadas no curso da pesquisa.

A terminologia aqui reunida e organizada segundo princípios ontológicos contribui, ainda que modestamente, para o assunto estudado, a preservação de registros digitais á medida em que se apreende a metodologia utilizada mais do que os resultados desse momento. Para tanto se busca fundamentação teórica na arquivologia, preservação de registros digitais e em segundo plano, nos princípios ontológicos, sempre presente no contexto desta pesquisa. Estudaram-se importantes iniciativas conceituadas pela comunidade acadêmica internacional com ênfase para a *OAIS* que se configurou em uma norma *ISO* o que a torna um padrão internacional com possibilidade de reconhecimento internacional.

Partindo-se disso conseguiu-se chegar a um conjunto de campos que se espera, possa contribuir de maneira modesta para novos estudos de preservação digital. Espera-se que o fato de estar embasada em iniciativas oriundas de instituições de grande renome pode garantir sua confiabilidade como trabalho científico e contribuir ainda para que exista uma concordância dos termos listados em língua portuguesa com os padrões mundiais o que contribuiria de certa maneira para a interoperabilidade do sistema.

A terminologia pode representar um universo particular da preservação digital em língua portuguesa o que irá indicar um caminho para diversas instituições nessa complexa tarefa. Definiram-se ali não apenas termos, mas contextos, classes e um universo criado especialmente para a preservação digital. Nesse sentido não pode ser considerada como um fim em si mesma para a preservação de documentos digitais, mas uma referência de criação de instrumentos de classificação pós coordenados que englobem as diversas facetas técnicas, arquivísticas, administrativas dentre outras.

O presente trabalho contribuiu com dados concretos sobre diversas iniciativas para a preservação digital, além do uso dos fundamentos da arquivística como ferramenta de apoio nessa prática. Embora a ênfase esteja em documentos arquivísticos, a pesquisa pode ser útil para as diversas áreas do conhecimento, especialmente para as organizações que precisam manter registros em meio digital por um longo prazo como parte de suas atribuições (por exemplo, bibliotecas digitais, museus e arquivos). Espera-se poder contribuir, ainda que modestamente, para futuros estudos de tópicos importantes para a preservação digital, como requisitos e estratégias. Acredita-se que possa auxiliar também usuários na preservação de seus registros pessoais.

Como resultado final e mais importante, obteve-se um modelo preliminar com a definição de um conjunto fundamental de metadados, além de relações entre os elementos de metadados, contribuindo assim no processo de elaboração de requisitos para preservação digital.

Mais importante que a estrutura em si mesmo é a iniciativa de utilização de princípios ontológicos na arquivística. É notório que a informação digital é parte do dia a dia das pessoas e organizações e demandas arquivísticas relacionadas a esse tipo de suporte já são uma realidade que não deve ser ignorada. Nesse sentido, princípios ontológicos para organização da informação representam uma alternativa na criação de Sistemas de Recuperação da Informação. Ontologias, se bem utilizadas, representam um avanço na pós coordenação da classificação que ganha ênfase nos arquivos digitais já que a busca se torna mais refinada e ampla dentro desse contexto. Espera-se assim que a presente iniciativa, possa servir de inspiração e de ponto de partida para outras.

REFERÊNCIAS

- ALMEIDA, Maurício Barcellos. **Um modelo baseado em ontologias para representação da memória organizacional**. 2006. 345 f. Tese (Doutorado) - Ufmg, Belo Horizonte, 2006.
- AMORIN, Eliane Dutra; LOPES, Carlos Eduardo Rodrigues; VALLE JÚNIOR, Eduardo Alves do. **Introdução à preservação de acervos digitais**. Belo Horizonte: Secretaria de Estado de Cultura, 2005. 42 p. (Cadernos Técnicos).
- ARANTES, Nelio. **Sistemas de Gestão Empresarial: Conceitos permanentes na administração de empresa válidas**. 2 ed. São Paulo: Atlas, 1998.
- BEARMAN, D.; SOCHATS, K. Functional Requirements for Evidence in Recordkeeping: The Pittsburgh Project (2002). Available from Internet: <<http://www.archimuse.com/papers/nhprc/BACartic.html>>. Access: 20 Jan. 2007.
- BELLOTO, Heloisa Liberalli. **Arquivos Permanentes: Tratamento Documental**. 4 ed. Rio de Janeiro: FGV, 2006. 320p.
- BRASIL.CONARQ. Dicionário brasileiro de Terminologia Arquivística. disponível em: <www.conarq.gov.br/>. Acessado em 15/07/2008.
- BRASIL.CONARQ. Norma Brasileira de Descrição Arquivística. disponível em: <<http://www.conarq.arquivonacional.gov.br/media/publicacoes/nobrade.pdf>>. Acesso em 21/07-2008.
- BRASIL. Lei No 8.159 de Janeiro de 1991. Dispõe sobre a política nacional de arquivos públicos e privados e da outras providências. Diário Oficial da União. Brasília, 9 jan. 1991.
- BRASIL.Presidência da República. Infra-estrutura de Chaves Públicas Brasileiras. Disponível em: <<https://www.icpbrasil.gov.br/>>. Acesso em: 29 de jul. 2007.
- BRASIL.Presidência da República.Resolução No 40, de abril de 2006. Aprova a versão 2.0 dos critérios e procedimentos para credenciamento das entidades integrantes da ICP-Brasil. Disponível em: <http://www.itl.twiki/.../resolu_0_40_de_18_04_2006.pdf>. Acesso em 27 jul. 2007.
- BULLOCK, Alison. **Preservation of Digital information: Issues and Current Status**. Ottawa: National Library of Canada, April 22 1999. Disponível em: <<http://epe.lac-bac.gc.ca/100/202/301/netnotes/netnotes-h/notes60.htm>>. Acesso em: 06 out. 2008.
- CASTRO, Astreá de Moraes; CASTRO, Andresa de Moraes; GASPARIAN, Danuza de Moraes. **ARQUIVOS: Físicos e digitais**. Brasília:Thesaurus, 2007. 192 p.
- CONSELHO INTERNACIONAL DE ARQUIVOS. Comitê de arquivos correntes em ambiente eletrônico. Documentos de arquivo eletrônicos: manual para arquivistas. ICA, Estudo nº 16. Disponível em: <<http://www.ica.org/biblio.php?pdocid=285>> Acesso em: 8 ago 2006.
- Constantopoulos, P., & Dritsou, V. (2007). An ontological model for digitalpreservation. *International Symposium in Digital Curation DigCCur2007, Chapel Hill, NC, April 2007*.
- CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS). **Reference model for an Open Archival Information System (OAIS)**; recommendation. Washington, 2002. 148p. Available from: <<http://www.ccsds.org/documents/650x0b1.pdf>>. Access: 14 Sept. 2008.

CORNWELL AFFILIATES PLC. *Electronic documents and record management*. Disponível em: <www.cornwell.co.uk e <http://www.cornwell.co.uk/moreq.html>>. Data de acesso: 01 ago.2009.

DCMI. **Dublin Core Metadata initiative**. Disponível em: <<http://dublincore.org/>>. Acesso em: 12 jun. 2010.

DEEGAN, Marilyn; SIMON, Tanner. Key Issues in digital preservation. In: DEEGAN, Marilyn; SIMON, Tanner. **Digital Preservation**. 1 ed. London:Facet Publishing, 2006. p. 01-31.

Dervin, B. (1994). Information <--> democracy: An examination of underlying assumptions. *Journal of the American Society for Information Science*, 45 (6), 369-385. (ERIC Document Reproduction Service No. EJ 488 232). Reprinted in: B. Dervin & L. Foreman-Wernet (with E. Lauterbach) (Eds.). (2003). *Sense-Making Methodology reader: Selected writings of Brenda Dervin* (pp. 73-100). Cresskill, NJ: Hampton Press.

DUBIN, David; PLUTCHAK, Joel. Metadata Enrichment for Digital Preservation. **Extreme Markup Languages**, Montreal, n. , p.1-11, 11 jul. 2006. Disponível em: <<http://www.digitalpreservation.gov/partners/resources/pubs/docs/EML2006Dubin01.pdf>>. Acesso em: 12 maio 2009.

FOUCAULT, Michel (Org.). **A microfísica do poder**. 2ª ed. Brasília: Graal, 2008. Organizado por Roberto Machado.

Governo Eletrônico. 2000f. Disponível na Internet em: <<http://www.governoeletronico.gov.br>>. Acesso 10 de jul. 2008.

HEDSTROM, Margaret. DIGITAL PRESERVATION: a time bomb for digital libraries. **Language Resources and Evaluation**, Springer (Netherlands), v. 31, n.3, p. 189 – 202, May. 1997.

INTERNATIONAL RESEARCH ON PERMANENT AUTHENTIC RECORDS IN ELETRÔNIC SYSTEMS. PRESERVATION TASK FORCE. **A MODEL OF THE PRESERVATION FUNCTION**. Available from Internet: <http://www.interpares.org/book/interpares_book_n_app05i.pdf>. Accessm: 02 Jun. 2008.

IKEMATU, Ricardo Shoití. 2001. **Gestão de Metadados**: sua evolução na tecnologia da informação. In: *Datagrama zero – Revista de Ciência da Informação*, 2(6), IASI – Instituto Brasileiro de Adaptação e Inserção na Sociedade da Informação.

International Concil of Museums. **CIDOC conceptual Reference Model**. Available from Internet: <<http://cidoc.ics.forth.gr>>. Access 10 Nov. 2008.

LAKATOS, E. M.; MARCONI, M. A. *Fundamentos de metodologia científica*. 3ed. São Paulo: Atlas, 1991.

LAZINGER, Susan S. Digital preservation and metadata: history, theory, practice. Englewood: Libraries Unlimited, 2001. 359p.

LE GOFF, J. In: _____. **História e Memória**. São Paulo: Ed. UNICAMP, 1984a.

LIBRARY OF CONGRESS (Estados Unidos). **METS**: Metadata Encoding & Transmition Standard. Disponível em: <<http://www.loc.gov/standards/mets/>>. Acesso em: 20 abr. 2009

LOPES, Ana Suely P. A GESTÃO DOCUMENTAL NAS ENTIDADES NACIONAIS DO SISTEMA INDÚSTRIA: Desafios e Soluções encontradas. Disponível em: www.asocarchi.cl/DOCS.39.PDF. Acessado em 21/07/2008.

LUSENET, Yola de. Towards Open Access in Europe. Paper presented at the seminar 'Open Access. Vom Prinzip zur Umsetzung/Du principe à la mise en pratique', Swiss Academy of Humanities and Social Sciences, 1 March 2007, Bern. Disponível em: <<http://www.knaw.nl/ecpa/publications.html>>. Acesso em: 01 fev. 2010.

MALHEIRO, da Silva Armando. **A gestão da informação arquivística e suas repercussões na produção do conhecimento científico.** [2001].

Minas Gerais. Deliberação No 2.392 de abril de 2001. Dispõe sobre o processo de digitalização de documentos na Assembléia Legislativa. Disponível em: <<http://hera.almg.gov.br/cgibin/...>>. Acesso em: 29 jul d 2008.

NATIONAL LIBRARY OF AUSTRALIA (Australia). **PADI: Preserving Access to Digital Information.** Disponível em: <<http://www.nla.gov.au/padi/metafiles/resources/42.html>>. Acesso em: 18 jun. 2009.

NEGREIROS, Leandro Ribeiro. *Sistemas eletrônicos de gerenciamento de documentos arquivísticos: um questionário para seleção avaliação e aplicação.* 2007. 248 f. Mestrado em Ciência da informação – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2007.

NATIONAL LIBRARY OF AUSTRALIA. The NLA Preservation Metadata Working Group. [1999]. Disponível em: <http://www.nla.gov.au/preserve/pmeta.html>. acesso em: 02 de Nov 2008.

PDF/A COMPETENCE CENTER (Alemanha). **PDF / A Conference Center.** Disponível em: <http://74.125.93.132/search?q=cache:http://www.pdfa.org/>>. Acesso em: 26 maio 2009.

SHELLENBERG, T. R. **Arquivos modernos: princípios e técnicas.** Tradução Nilza Teixeira Soares. 2ª ed. Rio de Janeiro: Editora FGV, 2002. 388p. Título original: Modern archives: principles and techniques (Chicago, USA, The University of Chicago Press, 1956)

THOMAZ, Kátia de Padua. *A preservação de documentos eletrônicos de caráter arquivístico.* 2004. 388f. Doutorado em Ciência da Informação – Escola de Ciência da Informação. Belo Horizonte. 2004.

UNITED KINGDOM. National Archives. Uk Government. Digital Preservation Guidance Note: 2: Selecting Storage Media for Long-Term Preservation. Preservation guide of National Archives of UK. Disponível em: <<http://www.nationalarchives.gov.uk/documents/selecting-storage-media.pdf>>. Acesso em: 05 mar. 2009.

UNIVERSITY OF CALIFORNIA, SCHOOL OF LIBRARY, ARCHIVAL AND INFORMATION STUDIES. *The long term preservation of authentic electronic records: findings of the inter pares project.* [2001]. Disponível em: <http://www.interpares.org/book/index.cfm>. acesso em: 23 de jun. 2008.

UNIVERSITY OF CALIFORNIA, SCHOOL OF INFORMATION MANAGEMENT AND SYSTEMS. How Much Information?. 18 out. 2000. Disponível em <<http://www2.sims.berkeley.edu/research/projects/how-much-info/summary.html#intro>>. Acesso em: 05 out. 2008

W.B. Croft, S. Cronen-Townsend, and V. Larvrenko. Relevance Feedback And Personalization: A Language Modeling Perspective. *Delos Workshop: Personalisation And Recommender Systems In Digital Libraries,* 2001.

Anexo 01

| Núm | Grupo | Definição Geral | Termos representativo | Especificações | Definições Específicas |
|-----|---------------------------------|---|--|---|--|
| 1 | Registration (Inscrição) | Consiste no grupo identificador do registro, nesse caso a transação comercial subdivide-se em dois termos | a) Metadados de Identificação do Registro (Não repetitivo) | Declaração dos Registros (Obrigatório) | Um <i>Bit Stream</i> que identifica que aquele dado é um registro. Pode ser consultado sem a abertura do registro, mas caso documento a que se refere seja aberto ele perde o valor. |
| | | | | Identificador do domínio da transação (Obrigatório) | Identifica o domínio que originou o registro: a operação e a organização responsável |
| | | | | Identificação de instância de Transação (Obrigatório) | Identifica uma instância de de operação com hora e data de criação do registro |
| | | | b) Metadados de Informações sobre Conteúdo | Conteúdo de Descrição Padrão (Opcional, exceto nos casos de definição de privacidade) | Identifica as normas de descrição do conteúdo, especialmente no que diz respeito a privacidade do documento |
| | | | | Contexto de Descrição (Opcional) | Fornece os termos utilizados pelo criador do registro para descreve-lo |
| | | | | Registro da Linguagem Natural Utilizada (Opcional) | Identifica o idioma em que o registro foi criado |

| Núm | Grupo | Definição Geral | Termos representativo | Especificações | Definições Específicas |
|-----|--|---|--------------------------------------|---|--|
| 2 | Terms & Conditions (Termos e Condições) | Define controles de acesso, utilização e a disposição de um registro. Identifica restrições de utilização e acesso. Subdivide-se em quatro sub campos | a) Metadados de Status de restrições | Status de Direito de Acesso (obrigatório) | Define se há e quais são as restrições de acesso do registro |
| | | | | Status de Direito de Uso | Define se há e quais são as restrições de direito de uso do registro |
| | | | b) Metadados de Condições de Acesso | Normas de Condições para Acesso | Identifica Condições que devem ser satisfeitas para acesso ao registro como: pagamentos, permissões, prova de identidade e outras restrições relacionadas com o acesso |
| | | | | Termos para Acesso | Define as condições de acesso de uma maneira de seja reconhecida pelo receptor |

| Núm | Grupo | Definição Geral | Termos representativo | Especificações | Definições Específicas |
|-----|--|---|----------------------------------|------------------------------|---|
| 2 | Terms & Conditions (Termos e Condições) | Define controles de acesso, utilização e a disposição de um registro. Identifica restrições de utilização e acesso. Subdivide-se em quatro sub campos | c) Metadados de Condições de Uso | Chaves para normas de uso | Identifica as chaves (resolvers) que devem ser satisfeitas. O usuário preenche as condições para utilização e o sistema é notificado sobre as formas de impor tais restrições. |
| | | | | Citação de Uso | Informações textuais fornidas pelo criador ou proprietário detalhando limitações a utilização |
| | | | | Regra de Leitura do Registro | Identifica o tipo de leitura permitido a usuários que não aqueles diretamente envolvidos e autorizados dentro da transação. Pode liberar a leitura por algoritmos ou requerer a intervenção humana. |
| | | | | Termos de Licença | Permitem a utilização do registro de acordo com diretrizes estabelecidas em licenças, se for esse o caso. |

| | | | | | |
|---|--|---|---|---|--|
| 2 | Terms & Conditions (Termos e Condições) | Define controles de acesso, utilização e a disposição de um registro. Identifica restrições de utilização e acesso. Subdivide-se em quatro sub campos | d) Metadados de políticas de informação | Autoridade de Remoção | identifica qual autoridade pode remover um registro ou parte dele do sistema. Estabelecido no momento da criação |
| | | | | Citação de políticas de Temporalidade | Composta de informações textuais para identificar a organização interna da política para temporalidade |
| | | | | Autoridade para regulamentação de temporalidade | Textos legais ou regulamentos institucionais que regem a temporalidade |
| | | | | Autoridade Externa de Temporalidade | Informações sobre o órgão que possui jurisdição sobre as leis ou regulamentos que regem a temporalidade |
| | | | | Período Final da Temporalidade | Indica a data em que o registro deve ser retirado do sistema. Deve conter citações a política de regulamentação e autoridade |
| | | | | Código de Disposição | Indica os métodos que se basearam a disposição final do registro |

| Núm | Grupo | Definição Geral | Termos representativo | Especificações | Definições Específicas |
|-----|---|---|---|----------------|--|
| 3 | Camada Estrutural (Structural Layer) | Consiste em metadados sobre a estrutura dos dados projetados para permitir que o registro se mantenha autêntico ao longo do tempo e que possa migrar para novos softwares e Hardwares, se necessário | a) Metadados de Identificação de arquivos individuais | ID do Arquivo | Identifica cada arquivo que compõem o registro dando condições para o sistema compor um dossiê de vários arquivos que formam um conjunto |

| Núm | Grupo | Definição Geral | Termos representativo | Especificações | Definições Específicas |
|-----|--------------------------------------|--|---|--|--|
| 3 | Camada Estrutural (Structural Layer) | Consiste em metadados sobre a estrutura dos dados projetados para permitir que o registro se mantenha autêntico ao longo do tempo e que possa migrar para novos softwares e Hardwares, se necessário | b) Metadados de Codificação de Arquivos | Propriedade do Arquivo | identifica a propriedade de cada arquivo (Texto, numérico, imagem, som, geográfico) |
| | | | | Códigos de representação de Dados do Arquivo | Identifica as normas de codificação de dados utilizadas pelo arquivo (ASCII, UNICODE, ASN, outros) |
| | | | | Códigos de Dados | Indica a forma como os dados são codificados e quais métodos estão sendo utilizados, por exemplo dados vetoriais podem ser topológicos |
| | | | | Método de Compressão de Dados | identifica o método de compressão de dados, se for esse o caso |
| | | | | Métodos de Criptografia de Dados | Identifica os algoritmos usados pelo autor para codificar os dados |

| Núm | Grupo | Definição Geral | Termos representativo | Especificações | Definições Específicas |
|-----|--------------------------------------|--|--|-------------------------------------|---|
| 3 | Camada Estrutural (Structural Layer) | Consiste em metadados sobre a estrutura dos dados projetados para permitir que o registro se mantenha autêntico ao longo do tempo e que possa migrar para novos softwares e Hardwares, se necessário | c) Metadados de Aparência dos Arquivos do Registro | Dependência de Aplicações | Indica de quais aplicações o registro é dependente |
| | | | | Dependência de Ambiente de Software | Indica a dependência de Software do registro incluindo o sistema operacionais e outros aplicativos, se for o caso. |
| | | | | Dependência de Hardware | Indica a dependência de Hardware do registro incluindo número do modelo, configuração, produção e informações (tais como impressoras ou leitores exigidos) |
| | | | | Saída do Registro | Identifica os procedimentos necessários para que o registro seja exibido, impresso ou outra forma que tenha sido representado no momento da criação |
| | | | | Padrão de Representação do Registro | Identifica qualquer norma aplicada ao processo que afeta a forma como o arquivo é apresentado (SGML, postScript, outras). Deve-se indicar a versão que foi utilizada. |

| Núm | Grupo | Definição Geral | Termos representativo | Especificações | Definições Específicas |
|-----|--------------------------------------|--|---------------------------------------|---|---|
| 3 | Camada Estrutural (Structural Layer) | Consiste em metadados sobre a estrutura dos dados projetados para permitir que o registro se mantenha autêntico ao longo do tempo e que possa migrar para novos softwares e Hardwares, se necessário | d) Metadados de Aparência do registro | Regras Padrão para Ligação entre arquivos | Identifica as regras ou normas necessárias a fim de permitir as ligações (links) necessárias entre os arquivos que compõem o registro |
| | | | | Dicionários de Aplicações | É o dicionário da base de dados. |
| | | | | Delimitadores / Rótulos | Consiste nos delimitadores e rótulos utilizados nos dados ou nas regras |
| | | | | Quadros de Busca de Dados e Valores | Consiste em um "arquivo autoridade" que contenha os valores dos códigos utilizados pelo registro e suas regras de utilização |
| | | | e) Metadados de estrutura do Conteúdo | Visualização de Dados na Criação | Identifica a forma como o aplicativo visualizou o registro no momento de sua criação. |
| | | | | Relacionamento de Versões | Consistem em qualquer identificador de registro de versões anteriores |
| | | | | Identificação de Registro de Relacionamento Comercial | Identifica o registro como pertencente a um grupo com fins comerciais dentro de um conjunto global de registros |
| | | | | Relacionamento Dinâmico | Identifica dados de outros registros que são exigidos para a complementação de valores. Define-se uma relação em que o arquivo não pode ser aberto se o conteúdo essencial, oriundo de outros arquivos não estiver disponível |

| Núm | Grupo | Definição Geral | Termos representativo | Especificações | Definições Específicas |
|-----|--------------------------------------|--|------------------------|---|--|
| 3 | Camada Estrutural (Structural Layer) | Consiste em metadados sobre a estrutura dos dados projetados para permitir que o registro se mantenha autêntico ao longo do tempo e que possa migrar para novos softwares e Hardwares, se necessário | f) Metadados de Origem | Origem / Fonte dos dados | Identifica a origem dos dados |
| | | | | Documentação de Data-Source | Consiste na documentação necessária para criar o registro - contém informações sobre o processamento dos dados |
| | | | | Tipo de Instrumento de Captura de Dados | Identifica o tipo de documento utilizado para capturar os dados (gravação sonora, temperatura, outros) e do instrumento específico (fabricante, número do modelo, etc) |
| | | | | Definições do Instrumento de Captura de Dados | Identifica as configurações do instrumento no momento da captura dos dados como calibração, por exemplo |
| | | | | Qualidade dos Dados | Identifica o grau de confiabilidade dos dados gerados pela fonte |
| | | | | | |

| Núm | Grupo | Definição Geral | Termos representativo | Especificações | Definições Específicas |
|-----|---------------------------------------|--|--|--------------------------------------|--|
| 4 | Contextual Layer (Camada de Contexto) | Identifica a Proveniência, ou seja a pessoa, organização ou sistema responsável pela geração do registro | Transaction Context metadata | Identificação do Originador | identifica a pessoa, organização ou sistema que iniciou a transação |
| | | | | Identificação do Destinatário | Identifica a organização / pessoa que recebeu a informação |
| | | | Metadados de Responsabilidade Geral | Organização de Origem | Identifica a unidade organizacional envolvida na transação |
| | | | Metadados de responsabilidade do sistema | Sistema de auditoria responsável | Citação da versão mais recente do sistema e das operações de auditoria interna |
| | | | | Auditoria executada no sistema | Citações mais recentes sobre auditorias internas no sistema |
| | | | | Consistência do Sistema de auditoria | Citação mais recente sobre as operações de auditoria interna que garantem a consistência desse sistema |

| Núm | Grupo | Definição Geral | Termos representativo | Especificações | Definições Específicas |
|-----|--|--|--|---|---|
| 5 | Content Layer (Camada de conteúdo) | Contém os dados reais contidos na transação | Conteúdo (traz o conteúdo do Registro) | Conteúdos Criados | Traz o conteúdo criado pela transação |
| | | | | Conteúdo Incorporado | Traz os dados que não fazem parte do registro original, mas foram incorporados os mesmo. |
| Núm | Grupo | Definição Geral | Termos representativo | Especificações | Definições Específicas |
| 6 | Use History Layer (Camada de história de uso) | Documentos importantes criados depois do registro. Incluem termos de indexação, Gravação, destruição, uso. Essa camada deve ficar sempre no fim do registro para permitir gravações sem que seja necessário abrir o registro | Metadados de História de Uso | Tipo de uso | identifica como os dados foram utilizados: visualizados, copiados, editado, arquivado, indexado, classificado, eliminados, etc. Deve-se observar a definição do usos permitidos pelo sistema. |
| | | | | Data e Hora do Uso | Identifica a data e a hora de uso dos dados |
| | | | | Identificação do usuário que fez acesso aos dados | identifica qual usuário fez acesso aos dados |
| | | | | Consequencias probatórias do uso | identifica o impacto do uso sobre aquele registro. |

Figura 5: Tabela de dados da Metadata Requirements Project

Anexo 02

Cabecalho METS

Nome da Iniciativa de Origem: METS

Termo Adotado: Identificação

Definição: Classe

Observações: A classe identificação refere-se a os dados que individualizam determinado objeto informacional, instituição ou outra entidade que necessite ser descrita durante os trabalhos.

-

Nome da Iniciativa de Origem: -

Termo Adotado: Objeto

Definição: Identifica uma sub-classe ou uma supra-classe; identifica-se se trata-se de um registro ou de uma entidade produtora de registros

Observações:

URI (Uniform Resource identifier)

Nome da Iniciativa de Origem: Dublin Core

Termo Adotado: Código Identificador do Objeto

Definição: Código Referente ao objeto.

Observações: Um código alfanumérico e único que será o identificador do documento

Identificador do domínio da transação

Nome da Iniciativa de Origem: Metadata Requirements Project

Termo Adotado: Código identificador do Domínio

Definição: Atribui um código para o domínio que enviou o documento

Observações: Esse código é especialmente válido para o caso de transações de negócios e identifica o remetente do objeto. Pode ser definido pelo IP da máquina que o enviou, por exemplo

Label

Nome da Iniciativa de Origem: Dublin Core

Termo Adotado: Etiqueta

Definição: Um rótulo, legível por linguagem humana, atribuído ao termo

Observações: Equivale a um título do objeto.

Definition

Nome da Iniciativa de Origem: Dublin Core

Termo Adotado: Definição

Definição: Uma definição que represente a natureza do termo; o conceito essencial

Observações:

Serial Number

Nome da Iniciativa de Origem: OAIS

Termo Adotado: Número Serial

Definição: Número seqüencial que identifica o documento ou pacote de informações

Observações: Esse número deve ser gerado automaticamente por sistemas informatizados para criar um número de série para cada registro

Persistent Identifier

Nome da Iniciativa de Origem: OAIS / Preservation Metadata for Digital Collections Initiative

Termo Adotado: URL Persistente

Definição: Identifica o *Hiperlink* utilizado para acesso ao objeto

Observações:

Version Number / Version

Nome da Iniciativa de Origem: OAIS / Dublin Core

Termo Adotado: Versão do Objeto

Definição: Identifica a versão do objeto

Observações: Todo objeto digital está sujeito a mudanças, logo é fundamental que se registre em que versão aquele objeto se encontra

Código de Referência / Label/ Object identifier

Nome da Iniciativa de Origem: NOBRADE / Dublin Core / OAIS

Termo Adotado: Código Identificador do objeto

Definição: Identifica o código atribuído ao objeto definido

Observações: Objeto livremente adaptado

Título / Nome / Name

Nome da Iniciativa de Origem: NOBRADE / Dublin Core / OAIS

Termo Adotado: Título do Objeto

Definição: Define nominalmente o objeto

Observações: Código livremente adaptado

-

Nome da Iniciativa de Origem: -

Termo Adotado: Título do Domínio

Definição: Define, nominalmente, o domínio

Observações: Objeto livremente adaptado. Registra o nome do domínio que pode ser uma máquina, um site, empresa ou pessoa física.

Resource Description / Metadados Descritivos

Nome da Iniciativa de Origem: CEDARS / METS

Termo Adotado: Descrição

Definição: metadado que descreve o objeto

Observações: Nessa sub classe podem ser incluídos metadados que descrevam o objeto

Data das descrições

Nome da Iniciativa de Origem: NOBRADE

Termo Adotado: Data da Descrição

Definição: Data em que a descrição foi preparada ou revisada.

Observações: Esse campo deve incluir datas de maneira cumulativa mantendo o histórico de alterações de descrição do objeto.

Bibliographic Description

Nome da Iniciativa de Origem: OAIS

Termo Adotado: Referência

Definição: Referência do objeto

Observações: Sempre que possível fazer a referência do objeto segundo os padrões vigentes

Definition / File Description

Nome da Iniciativa de Origem: Dublin Core / Preservation Metadata for Digital Collections Initiative

Termo Adotado: Definição do Objeto

Definição: Cria uma definição do objeto

Observações: Nesse metadados inclui-se uma definição / descrição do objeto

-

Nome da Iniciativa de Origem: -

Termo Adotado: Tipo do Objeto

Definição: Define o tipo de objeto

Observações: Define se o objeto é um arquivo, um pacote de dados ou outro tipo de informação

-

Nome da Iniciativa de Origem: -

Termo Adotado: Descrição do Tipo de Objeto

Definição: Descreve o Tipo de Objeto

Observações: Descreve o conteúdo do Tipo de Objeto se é um pacote de informação, por exemplo, o que contém.

Nota do Arquivista

Nome da Iniciativa de Origem: NOBRADE

Termo Adotado: Notas de Arquivista

Definição: Refere-se a notas técnicas que o arquivista ou profissional técnico responsável julgue necessárias.

Observações: Descreve as fontes utilizadas, códigos de referência de documentos utilizados, Nome e responsabilidade das pessoas envolvidas no trabalho.

Regras ou Convenções

Nome da Iniciativa de Origem: NOBRADE

Termo Adotado: Normas ou Convenções

Definição: Identifica as regras ou convenções em que se baseia a descrição do objeto

Observações: Segundo determinação da NOBRADE os objetos devem indicar as normas em que basearam suas descrições.

Data Object

Nome da Iniciativa de Origem: CEDARS

Termo Adotado: Data

Definição: Sub-classe que agrupa informações de data

Observações:

Data de produção / Date of Criation

Nome da Iniciativa de Origem: NOBRADE / Preservation Metadata for Digital Collections Initiative

Termo Adotado: Data de Criação do Objeto

Definição: Data em que o objeto foi criado

Observações: Código adaptado livremente

-

Nome da Iniciativa de Origem: -

Termo Adotado: Suporte

Definição: Sub-classe que agrupa dados sobre o suporte do objeto

Observações: Código Adaptado livremente

Dimensão e Suporte

Nome da Iniciativa de Origem: NOBRADE

Termo Adotado: Tipo de Suporte

Definição: Tipo de Suporte em que o objeto está acondicionado

Observações: Como sabido todo objeto digital depende de um suporte externo para ser acondicionado. Esse campo define nominalmente esse suporte

Dimensão do Suporte

Nome da Iniciativa de Origem: NOBRADE

Termo Adotado: Dimensão do Suporte

Definição: Dimensão do suporte utilizado no armazenamento

Observações: Dimensão do suporte é algo importante para o planejamento do acondicionamento atual e futuro.

Digital Collection management / Metadados Administrativos

Nome da Iniciativa de Origem: CEDARS / METS

Termo Adotado: Gerenciamento de Coleções Digitais

Definição: Refere-se ao processo de gerenciamento das coleções.

Observações:

Producer / Record Creator

Nome da Iniciativa de Origem: OAIS / Preservation Metadata for Digital Collections Initiative

Termo Adotado: Produtor

Definição: Classe de Informações sobre o produtor / autor / criador do objeto

Observações: Agrupa os metadados referentes a produção do documento

Archival Information Packages

Nome da Iniciativa de Origem: OAIS

Termo Adotado: Armazenamento

Definição: Classe de informações sobre o sistema de guarda do(s) objeto(s)

Observações: Essa classe agrupa os metadados referentes ao sistema de guarda e / ou armazenamento dos objetos

-

Nome da Iniciativa de Origem: -

Termo Adotado: Informação de Armazenamento

Definição: Refere-se a informações sobre as condições de armazenamento do objeto

Observações:

Incorporação

Nome da Iniciativa de Origem: NOBRADE

Termo Adotado: Informação de Incorporação de Conteúdo

Definição: Refere-se a incorporação de novos conteúdo ao objeto.

Observações: Quando um conteúdo é submetido a um objeto previamente preservado este deve ser indicado para manter um histórico de alterações.

Institution with preservation responsibility

Nome da Iniciativa de Origem: Preservation Metadata for Digital Collections Initiative

Termo Adotado: Instituição Responsável

Definição: O nome da instituição que aceitou a responsabilidade pela preservação. Deverá incluir a data de início da aceitação e / ou os intervalos de datas de responsabilidade.

Observações:

Storage Information

Nome da Iniciativa de Origem: Preservation Metadata for Digital Collections Initiative

Termo Adotado: Informações de armazenamento

Definição: Define a capacidade de armazenamento do sistema e detalhes como formatos suportados.

Observações:

Condições de Uso / Condições de Acesso

Nome da Iniciativa de Origem: Metadata Requirements Project

Termo Adotado: Condições de Uso e Acesso

Definição: Sub classe da classe armazenamento. Reúne os termos relacionados com as condições de uso e Acesso aos Documentos

Observações:

Status de Direito de Acesso

Nome da Iniciativa de Origem: Metadata Requirements Project

Termo Adotado: Direito de Acesso

Definição: Refere-se ao direito de acesso a determinado objeto por parte de um grupo específico de usuários

Observações:

Instrumentos de Pesquisa

Nome da Iniciativa de Origem: NOBRADE

Termo Adotado: Instrumentos de pesquisa relacionados

Definição: Descreve instrumentos auxiliares de pesquisa relacionados com o objeto.

Observações: Muitas vezes instrumentos de pesquisa pós coordenados são utilizados para recuperar documentos. Esse campo relaciona quais os instrumentos estão ligados ao objeto.

Normas de Condições para acesso

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Normas de acesso

Definição: Define as condições de acesso ao objeto como pagamento, permissões, prova de identidade, dentre outras

Observações: Esse objeto relaciona-se com as políticas institucionais relacionadas com o objeto excluindo-se portanto as condições técnicas.

Termos para Acesso

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Termos de acesso

Definição: Define as condições técnicas para o acesso ao objeto.

Observações: Refere-se as especificações técnicas de acesso ao objeto. Deve ser adequada as condições de entendimento do sistema que ira receber os dados.

Access inhibtors

Nome da Iniciativa de Origem: Preservation Metadata for Digital Collections Initiative

Termo Adotado: Inibidores de Acesso

Definição: Define qualquer método utilizado para inibir o acesso que impacte nos procedimentos de preservação.

Observações: Muitas vezes os documentos digitais são protegidos com criptografia ou assinaturas digitais isso deve ser descrito para efeitos de preservação.

Termos de Licença

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Licença de uso

Definição: Define as permissões de uso do objeto de acordo com diretrizes estabelecidas em licenças

Observações: Em geral softwares de uso proprietário ou obras de que contém direitos autorais estabelecem condições de uso como restrições ao uso público, por exemplo.

Metadados de Políticas de Informação

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Políticas de Informação

Definição: Sub classe da classe armazenamento. Agrupa os termos relacionados com políticas de informação dentro do gerenciamento das coleções.

Observações: As políticas de informação devem ser definidas previamente segundo legislação vigente e normas da pessoa ou instituição responsável pelo gerenciamento dos objetos.

Archiving decision

Nome da Iniciativa de Origem: Preservation Metadata for Digital Collections Initiative

Termo Adotado: Decisões Arquivísticas

Definição: Reúne o conjunto de decisões tomadas pelo responsável técnico pelo arquivo que envolvem o objeto em questão.

Observações:

Action History

Nome da Iniciativa de Origem: CEDARS

Termo Adotado: Histórico de Ações

Definição: Reúne informações acerca do histórico de ações que forma tomadas de decisão afim de preservar o objeto.

Observações:

Policy Actions

Nome da Iniciativa de Origem: CEDARS

Termo Adotado: Histórico de Políticas

Definição: Reúne o histórico das políticas de informação relacionadas com o objeto.

Observações:

História Arquivística

Nome da Iniciativa de Origem: NOBRADE

Termo Adotado: Histórico Arquivístico

Definição: Reúne o histórico arquivístico do objeto.

Observações: Oferecer informações referenciais sistematizadas sobre a história da produção e armazenamento do objeto.

Autoridade de Remoção

Nome da Iniciativa de Origem: Metadata Requiriments Project

Termo Adotado: Autorização para Exclusão de Objeto

Definição: Metadado que remete a autoridade necessária para exclusão do objeto. Deve ser determinada sempre pelo administrador do sistema

Observações: Sempre que um objeto for criado, ou transferido para a área de armazenamento deve-se definir quem tem autorização para excluí-lo, em geral essa autorização refere-se ao produtor e / ou administrador do sistema, mas deve obedecer às políticas de informação previamente definidas

Citação de Políticas de Temporalidade / Avaliação, Eliminação e Temporalidade

Nome da Iniciativa de Origem: Metadata Requeriments Project / NOBRADE

Termo Adotado: Políticas de Temporalidade

Definição: Metadado que descreve as políticas de temporalidade do objeto segundo os preceitos arquivísticos das três idades

Observações: Os objetos devem ter um tempo de guarda, assim como normas para recolhimento e transferência para outros ambientes ou suportes definidos pelas políticas de informação da instituição.

Autoridade para regulamentação da autoridade

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Normas / Legislações para Temporalidade

Definição: Cita as normas e legislações ou autoridades que justifiquem a adoção de determinada temporalidade

Observações: Em geral a temporalidade se da por lei ou normas internas de cada instituição. Caso não existam pode ser determinada pela autoridade competente legalmente constituída.

Consumer

Nome da Iniciativa de Origem: OAIS

Termo Adotado: Disseminação

Definição: Classe de informações sobre o sistema de disseminação dos objetos

Observações: Agrupa metadados referentes à disseminação, disponibilização dos objetos para o consumidor final.

Pontos de Acesso e indexação de Assuntos

Nome da Iniciativa de Origem: NOBRADE

Termo Adotado: Indexação de Assuntos

Definição: Registra termos para indexação do objeto

Observações:

Área de Notas

Nome da Iniciativa de Origem: NOBRADE

Termo Adotado: Notas

Definição: Classe para informações que não sejam contempladas nas áreas anteriores

Observações: Incluem-se informações relevantes que não sejam descritas nos tópicos anteriores.

Comment / Other / Notas Gerais

Nome da Iniciativa de Origem: Dublin Core / Preservation Metadata for Digital Collections Initiative / NOBRADE

Termo Adotado: Comentário

Definição: Informações adicionais sobre o Metadado ou classe e sua aplicação.

Observações: Inclui informações não contempladas em outros campos que possam ser úteis para definir o termo e sua aplicação.

References / Resource Description

Nome da Iniciativa de Origem: Dublin Core / CEDARS

Termo Adotado: Referências

Definição: Inclui as referências dos recursos utilizados na constituição do termo, se esse for o caso.

Observações:

Provenance

Nome da Iniciativa de Origem: OAIS

Termo Adotado: Proveniência

Definição: Essa classe agrupa campos referentes ao princípio da proveniência, base da arquivística.

Observações:

Procedência

Nome da Iniciativa de Origem: NOBRADE

Termo Adotado: Procedência

Definição: Identifica a origem imediata de aquisição ou transferência da unidade de descrição.

Observações:

Sistema de Arranjo

Nome da Iniciativa de Origem: NOBRADE

Termo Adotado: Sistema de Arranjo

Definição: Informa sobre o sistema de arranjo do objeto.

Observações: O sistema de arranjo é parte fundamental de um objeto arquivístico, pois remete diretamente ao princípio de proveniência.

Original Technical Environments

Nome da Iniciativa de Origem: CEDARS

Termo Adotado: Ambiente Técnico Original

Definição: Reúne metadados relacionados com o ambiente técnico onde o documento foi criado.

Observações:

Prerequisites

Nome da Iniciativa de Origem: CEDARS

Termo Adotado: Pré requisitos

Definição: Identifica os pré requisitos técnicos necessários para implantação do objeto.

Observações: Todo objeto digital necessita de pré requisitos técnicos de *hardware* e *Software* para ser instalado dentro de determinado ambiente. Esse metadado guarda esses requisitos.

Procedures

Nome da Iniciativa de Origem: CEDARS

Termo Adotado: Procedimentos

Definição: Identifica os procedimentos operacionais necessários para criação ou recebimento do objeto

Observações: Descreve os procedimentos da instituição ou relacionados com a transferência do objeto no momento de sua recepção por parte do responsável pela guarda.

Documentation

Nome da Iniciativa de Origem: CEDARS

Termo Adotado: Documentação

Definição: Identifica a documentação relacionada ao objeto na época de sua criação.

Observações: Identifica a documentação relacionada ao objeto no momento de sua criação ou de seu recolhimento por parte do responsável pela preservação.

Tecnichal Infrastructure of complex object

Nome da Iniciativa de Origem: Preservation Metadata for Digital Collections Initiative

Termo Adotado: Infra Estrutura Técnica de Objetos Complexos

Definição: Descreve a infra estrutura técnica de objetos complexos.

Observações: Alguns objetos como bases de dados e páginas da WEB possuem uma estrutura técnica complexa que deve ser descrita.

Processing History

Nome da Iniciativa de Origem: OAIS

Termo Adotado: Histórico do processamento

Definição: Refere-se ao histórico do processamento dos dados relacionados ao objeto

Observações: Esse campo é utilizado fundamentalmente em dados relacionados com software e aplicativos que tenham sido compilados ou que tenham sofrido algum tipo de conversão relacionada com o processamento.

Software interface Specification

Nome da Iniciativa de Origem: OAIS

Termo Adotado: Especificação de Interface de Software

Definição: Refere-se a interface de software utilizada na criação do objeto.

Observações: Define características do software: se é proprietário, qual o fabricante, a versão utilizada, que a interface de sistema operacional, dentre outros.

Custody history

Nome da Iniciativa de Origem: CEDARS

Termo Adotado: Histórico da Custódia do Objeto

Definição: Refere-se à custódia do objeto desde a sua criação até a data atual.

Observações: Esse dado é especialmente importante para arquivos. Muitas vezes uma secretaria tem suas atividades encerradas e o fundo referente é transferido para sua sucessora, nesse caso é importante registrar a mudança de custódia do acervo.

Metadata about digitalization process

Nome da Iniciativa de Origem: OAIS

Termo Adotado: Processo de Digitalização

Definição: Inclui as informações sobre o processo de digitalização.

Observações: Inclui metadados sobre o processo de digitalização. Tipo de equipamento, data, responsável dentre outros necessárias segundo o propósito.

Copyright / Rights Managment

Nome da Iniciativa de Origem: OAIS / CEDARS

Termo Adotado: Copyright

Definição: Classe que inclui metadados sobre os direitos autorais do objeto, se for esse o caso.

Observações:

Negotiation history

Nome da Iniciativa de Origem: CEDARS

Termo Adotado: Histórico de Negociação

Definição: Inclui o histórico das negociações de direitos autorais do objeto.

Observações: Esse termo, como muitos outros relacionados com preservação digital, se traduz em uma lista de auditoria com os acordos e negociações estabelecidas com o detentor dos direitos autorais.

Rights Information

Nome da Iniciativa de Origem: CEDARS

Termo Adotado: Informações de Direitos Autorais

Definição: Informações de cunho geral sobre os direitos autorais do objeto

Observações:

Publisher

Nome da Iniciativa de Origem: CEDARS

Termo Adotado: Editor

Definição: Informa o nome do editor / criador do objeto.

Observações:

Place of Publication

Nome da Iniciativa de Origem: CEDARS

Termo Adotado: Local de Publicação

Definição: Informa o local onde o objeto foi publicado em sua primeira edição.

Observações:

Rights Warning

Nome da Iniciativa de Origem: CEDARS

Termo Adotado: Legislação Relacionada

Definição: Informa a legislação vigente na época da criação do objeto.

Observações:

Contacts of Rights Holder

Nome da Iniciativa de Origem: CEDARS

Termo Adotado: Contatos do Detentor dos Direitos Autorais

Definição: Informa o histórico de contatos e negociação com o detentor dos direitos autorais.

Observações: Informa o histórico de contatos e negociação com o detentor dos direitos autorais.

Actions

Nome da Iniciativa de Origem: CEDARS

Termo Adotado: Ações

Definição: Reúne os metadados relacionados com as ações permitidas para o objeto.

Observações:

Permitted by Statute

Nome da Iniciativa de Origem: CEDARS

Termo Adotado: Permitidas pelo estatuto

Definição: Informa quais ações podem ser executadas para o objeto segundo o estatuto / regulamento da instituição que o preserva.

Observações:

Permitted by License

Nome da Iniciativa de Origem: CEDARS

Termo Adotado: Permitida pela Licença

Definição: Informa quais ações podem ser executadas para o objeto segundo a licença de uso do fabricante.

Observações:

Registration

Nome da Iniciativa de Origem: OAIS

Termo Adotado: Registro

Definição: Inclui informações sobre registro do objeto nas instâncias e órgãos competentes, se for esse o caso

Observações:

Context

Nome da Iniciativa de Origem: OAIS

Termo Adotado: Contexto

Definição: Essa classe agrupa campos referentes ao contexto geral em que o documento está inserido

Observações: Uma informação muito importante no âmbito da preservação e entender o contexto no qual o documento foi criado. Para tanto é necessária a criação de metadados nesse sentido.

Help File

Nome da Iniciativa de Origem: OAIS

Termo Adotado: Arquivo de Ajuda

Definição: Arquivo criado para fornecer ajuda na contextualização do objeto

Observações: Esse arquivo deve prestar ajuda ao usuário no entendimento do contexto de criação do objeto.

Existência de Localização dos Originais

Nome da Iniciativa de Origem: NOBRADE

Termo Adotado: Existência ou Localização dos Originais

Definição: Indica a localização ou inexistência dos originais

Observações: Em alguns casos os objetos podem ser cópias e nesse caso indica-se a localização do original. Isso remete fortemente ao contexto probatório do documento no ambiente digital.

Existência e Localização das cópias

Nome da Iniciativa de Origem: NOBRADE

Termo Adotado: Existência e Localização das cópias

Definição: Indica a Localização de cópias do documento

Observações: A distribuição de cópias do objeto deve ser informada nesse campo.

Unidades de Descrição relacionadas

Nome da Iniciativa de Origem: NOBRADE

Termo Adotado: Unidades de Descrição relacionadas

Definição: Identifica a existência de unidades de descrição relacionadas

Observações: Caso existam níveis e unidade de descrição que se relacionem de alguma maneira tais relacionamentos devem ser descritos nesse campo.

Quirks

Nome da Iniciativa de Origem: Preservation Metadata for Digital Collections Initiative

Termo Adotado: Funcionalidade / Aparência

Definição: Indica se houve mudanças na aparência ou perdas de funcionalidade

Observações: Durante os processo de migração podem haver perdas nas características de aparência do objeto ou mudanças na funcionalidade de alguns recurso.

Language / Idioma

Nome da Iniciativa de Origem: OAIS / NOBRADE

Termo Adotado: Idioma

Definição: Indica o idioma em que o objeto foi criado.

Observações: O idioma em que o objeto foi criado pode significar um dos conhecimentos básicos para entendimento do mesmo.

Fixity

Nome da Iniciativa de Origem: OAIS

Termo Adotado: Autenticidade

Definição: Essa classe agrupa metadados relacionados com a garantia de autenticidade do objeto

Observações:

Digital Signature

Nome da Iniciativa de Origem: OAIS

Termo Adotado: Assinatura Digital

Definição: Descreve o modelo utilizado para assinar o objeto digitalmente, caso seja esse o caso.

Observações: Existem diversos modelos de assinatura digital de objetos. O Brasil regulamenta o modelo oficial através da ICP-Brasil, previamente descrita nesse trabalho.

Authenticity Indicator

Nome da Iniciativa de Origem: OAIS e CEDARS

Termo Adotado: Indicador de Autenticidade

Definição: Indica qual o sistema ou certificação garante a autenticidade do objeto.

Observações:

Certificate

Nome da Iniciativa de Origem: OAIS

Termo Adotado: Certificado

Definição: Refere-se a certificados relacionados com os requisitos para autenticidade do objeto.

Observações: Esse termo pode incluir certificados de softwares, objetos, assinatura digitais e outros necessários para garantir que o documento seja aceito como autentico.

Structural Layer

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Dados Estruturais

Definição: Essa classe agrupa metadados relacionados com a estrutura de dados.

Observações: A estrutura de dados está relacionada com informações técnicas fundamentais dentro do processo de migração, autenticidade e controle dos registros ao longo do tempo.

ID do Arquivo

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Identificação do objeto Individual

Definição: Identifica de maneira individual cada objeto.

Observações: Além de identificar cada objeto esse número pode contar com uma identificação, previamente definida, que vincule o objeto a uma determinada classe ou dossiê.

Propriedade do Arquivo

Nome da Iniciativa de Origem: Metadata Requeriments Project / Preservation Metadata for Digital Collections Initiative

Termo Adotado: Tipo do objeto

Definição: Identifica a propriedade de cada objeto (Texto, Som, Imagem)

Observações: Essa categoria pode ser pré definida se assim for necessário ou viável dentro da política de informação pré definida. Podem existir mais uma categoria dentro do mesmo objeto como no caso das páginas WEB.

Códigos de Representação de Dados do Arquivo

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Código de representação

Definição: Identifica normas de codificação utilizadas pelo objeto e seu ambiente operacional.

Observações: Existem diversos códigos utilizados para compilar a linguagem de máquina para linguagem inteligível por seres humanos (ASCII – UNICODE – ASN – Outros)

Código de Dados

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Outros Códigos

Definição: Identifica outros códigos que por ventura sejam utilizados pelo objeto.

Observações: Refere-se a códigos necessários para o entendimento do contexto como dados vetoriais que se referem a estruturas topológicas, por exemplo. No futuro essas escalas podem ser diferentes e o entendimento pode ficar prejudicado.

Método de Compressão de Dados

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Método de Compressão

Definição: Identifica o método de compressão de dados, se for esse o caso.

Observações:

Métodos de Criptografia de Dados

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Método de criptografia

Definição: Identifica os algoritmos utilizados pelo autor se esse for o caso.

Observações: Pode referir-se a algoritmos ou classes de algoritmos.

Dependência de Aplicações

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Dependência de Aplicações

Definição: Identifica quais aplicações o registro depende para ser visualizado

Observações:

Dependência de Ambiente de Software

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Dependência de Ambiente de *Software*

Definição: Identifica de quais *software* o objeto é dependente.

Observações: Deve incluir aplicativos, sistemas operacionais e todos os demais programas necessários para visualizar e alterar, se for o caso, o objeto.

Know System Requeriments

Nome da Iniciativa de Origem: Preservation Metadata for Digital Collections Initiative

Termo Adotado: Requisitos de Sistema Conhecidos

Definição: Define os requisitos para que o sistema original funcione.

Observações: Pode incluir tanto requisitos de softwares auxiliares quando de Hardware. É importante indicar se os requisitos são obrigatórios.

Installation Requeriments

Nome da Iniciativa de Origem: Preservation Metadata for Digital Collections Initiative

Termo Adotado: Requisitos de Instalação

Definição: Define os requisitos para que o objeto seja instalado.

Observações: Define requisitos de instalação como plugins e condições de contrato e outras relacionadas.

De pendência de *Hardware* / Equipamentos e meios necessários para uso da unidade de descrição

Nome da Iniciativa de Origem: Metadata Requeriments Project / NOBRADE

Termo Adotado: Dependência de Hardware

Definição: Identifica de qual *Hardware* o objeto é dependente.

Observações: Deve fazer uma identificação completa de *hardware* incluindo número do modelo, configuração, produção e informação sobre impressoras, leitores ou outros necessários. Inclui ainda os meios para obtenção do equipamento, se for o caso.

Saída do Registro

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Condições técnicas de acesso

Definição: Identifica os procedimentos necessários para que o registro seja exibido, impresso ou outra forma que tenha sido prevista na criação.

Observações:

Condições de Acesso

Nome da Iniciativa de Origem: NOBRADE

Termo Adotado: Condições de Acesso

Definição: Fornecer condições legais ou estatutárias para acesso ao objeto

Observações: Refere-se a informações não técnicas de acesso ao objeto como legais ou estatutárias.

Condições de Reprodução

Nome da Iniciativa de Origem: NOBRADE

Termo Adotado: Condições de Reprodução

Definição: Identifica qualquer restrição para reprodução do objeto.

Observações: Esse campo determina se existem restrições para reprodução / cópia do objeto e quais usuários ou grupos estão autorizados a realizar a ação.

Ligações Estruturais

Nome da Iniciativa de Origem: METS

Termo Adotado: Ligações Estruturais

Definição: Permite registrar a existência de hiperlinks na estrutura do objeto.

Observações: Esse recurso é especialmente importante no caso de Sites.

Padrão de representação do Registro

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Padrão de Representação de objeto

Definição: Identifica qualquer norma aplicada ao processo que afete a forma como o objeto é apresentado (SGML, PostScript, XML, outras)

Observações: (SGML, PostScript, XML, outras) Deve-se indicar a versão que foi utilizada, fabricante e outras informações importantes para a recuperação do objeto.

Regras Padrão para Ligação entre arquivos

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Normas para ligação entre Hipertexto

Definição: Identifica Normas de acesso para Links que apontam para ambientes externos.

Observações: Em alguns casos os hiperlinks podem apontar para ambientes diferentes ou o acesso outros ambientes pode ser restrito pelos gestores.

Padrão de Intercâmbio entre arquivos

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Padrão de Intercâmbio entre objetos

Definição: Identifica normas ou regras necessárias para o intercâmbio entre arquivos de um mesmo dossiê ou registro

Observações: Inclui as versões apropriadas dentre outros dados importantes para manter o intercâmbio entre os diversos objetos.

Visualização de Dados na Criação

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Aspecto visual do objeto na criação

Definição: Identifica e mostra a forma como o objeto foi visualizado no momento de sua criação.

Observações: Pode incluir descrições de texto e imagens do objeto.

Relacionamento de Versões / Seção de Ficheiros

Nome da Iniciativa de Origem: Metadata Requeriments Project / METS

Termo Adotado: Identificador das versões Anteriores

Definição: Identifica as versões anteriores do objeto e cria ligação com a atual

Observações: Todo processo de migração deve ter a opção de ser revertido. Para tanto a ligação entre o objeto atual e suas versões anteriores é fundamental.

Identificação de Registro de Relacionamento Comercial

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Identificador de Registro de Transação Comercial

Definição: Identifica o registro como pertencendo a um grupo com fins comerciais dentro de um universo maior.

Observações: Grande parte da documentação digital, especialmente com fins probatórios e certificação está inserida dentro do contexto das transações comerciais. Nesse sentido é importante identificar quais transações estão relacionadas com registros comerciais.

Relacionamento Dinâmico / Mapa Estrutural

Nome da Iniciativa de Origem: Metadata Requeriments Project / METS

Termo Adotado: Relacionamento Estrutural

Definição: Identifica dados de outros objetos que são exigidos para a complementação de valores. Define-se uma relação entre os objetos para que o arquivo não seja aberto sem esses dados.

Observações: Algumas aplicações, especialmente que envolvem padrão dinâmicos de dados, como XML, podem exigir a presença de dados externos para complementação das informações a

serem exibidas. Esse campo de metadado inclui essa relação e a busca pelas referidas informações.

Origem Fonte dos Dados

Nome da Iniciativa de Origem: Verificar de quais iniciativas esses campos são oriundos.

Termo Adotado: Origem / Fonte dos dados

Definição: Identifica a origem dos dados determinado objeto.

Observações:

Documentação de Data-Source

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Data-Source

Definição: Indica a documentação necessária para a criação do registro

Observações: Contém informações sobre o processamento de dados

Tipo de Instrumento de Captura de Dados

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Instrumento de Captura de Dados

Definição: Identifica o(s) instrumento(s) utilizados na obtenção dos dados do registro.

Observações:

Definições do Instrumento de Captura de Dados

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Definição / Configuração do Instrumento de captura dos dados

Definição: Identifica o tipo e a configuração do instrumento no momento de captura dos dados.

Observações: Esse metadado identifica o tipo de recurso, modelo, marca e as configurações utilizadas no momento da captura . Um exemplo é uma máquina que captura uma fotografia digital.

Qualidade dos dados

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Qualidade da fonte de dados

Definição: Identifica a qualidade / confiabilidade da fonte de dados que gerou o objeto.

Observações: Algumas informações podem ser oriundas de fontes duvidosas como alguns sites da internet, por exemplo. Nesse caso deve-se apontar qual o grau de confiabilidade das informações.

Identificador de Recepção

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Identificador de Recepção

Definição: Identifica o cargo, pessoa e / ou sistema que recebeu o objeto, além da data e hora.

Observações:

Identificador de Origem

Nome da Iniciativa de Origem: Metadata Requeriments Project

Termo Adotado: Origem do Objeto

Definição: Identifica a pessoa, organização ou sistema que gerou aquele objeto.

Observações:

-

Nome da Iniciativa de Origem:

Termo Adotado: *Software* relacionados

Definição: Identifica os softwares relacionados com o objeto

Observações: Termo adaptado livremente

Related Information Objects / Relationships

Nome da Iniciativa de Origem: CEDARS / Preservation Metadata for Digital Collections Initiative

Termo Adotado: Objetos de Informação Relacionados

Definição: Indica outros objetos que estejam relacionados e sejam importantes para entendimento do contexto no momento de recebimento ou criação.

Observações:

Autorização

Nome da Iniciativa de Origem: Metadata Requirements Project

Termo Adotado: Identifica quem autorizou o ato que originou o objeto.

Definição: Todo documento arquivístico é oriundo de um ato. Esse campo identifica a Pessoa física ou jurídica que originou esse ato. No caso de transações comerciais identificam-se as partes que autorizaram o negócio.

Observações:

Sistema de Auditoria Responsável

Nome da Iniciativa de Origem: Metadata Requirements Project

Termo Adotado: Sistema de auditoria responsável

Definição: Identifica o sistema e o procedimento utilizado para auditoria no que diz respeito a autenticidade das informações.

Observações: Identifica qual o sistema de auditoria garantem a certificação do sistema que gerou o objeto. No caso do Brasil, por exemplo, os sistemas devem ser homologados por entidade credenciada pelo ICP-Brasil.

Sistema de Auditoria de Execução

Nome da Iniciativa de Origem: Metadata Requirements Project

Termo Adotado: Sistema de auditoria de execução

Definição: Identifica o sistema e o procedimento utilizado para auditoria no que diz respeito a implementação do sistema.

Observações: Refere-se a implantação efetiva do sistema e sua manutenção segundo os critérios determinados pela autoridade Certificadora.

Sistema de Auditoria

Nome da Iniciativa de Origem: Metadata Requirements Project

Termo Adotado: Sistema de Auditoria

Definição: Identifica o sistema e o procedimento utilizado para auditoria no que diz respeito a Consistência do sistema.

Observações: Além do sistema de auditoria e da manutenção do mesmo importante demonstrar a consistência do sistema no decorrer do processo já que o mesmo pode estar sujeito a falhas e depende de melhorias contínuas para alcançar e manter um padrão de qualidade.

Histórico de Usos

Nome da Iniciativa de Origem: Metadata Requirements Project

Termo Adotado: Sub Classe vinculada a classe de autenticidade que agrupa os metadados que registram o histórico de uso do objeto

Definição: Essa sub-classe foi criada com fins de agrupar dados relacionados com o histórico de uso. Embora não conste formalmente de nenhuma iniciativa ela agrupa dados que são oriundos e um ou mais sistemas de preservação.

Observações:

Tipo de Uso

Nome da Iniciativa de Origem: Metadata Requirements Project

Termo Adotado: Trilha de Auditoria de uso do Objeto

Definição: Identifica a maneira como o objeto foi utilizado.

Observações: Identifica as operações relativas ao objeto: Visualização, cópia, edição, transferência ou recolhimento, classificação, reclassificação, envio, eliminação

Data de Uso

Nome da Iniciativa de Origem: Metadata Requirements Project

Termo Adotado: Data de Uso

Definição: Identifica a data e a hora em que o objeto foi utilizado.

Observações: Também deve trabalhar no sistema de trilha de auditoria registrando todas as datas e horas em que o objeto foi acessado.

Usuário

Nome da Iniciativa de Origem: Metadata Requirements Project

Termo Adotado: Idêntica o usuário que utilizou o objeto

Definição: Deve fazer uma trilha de auditoria que conste todos os usuários que utilizaram o arquivo.

Observações: Deve fazer uma trilha de auditoria que conste todos os usuários que utilizaram o arquivo.

Conseqüência do uso para efeitos probatórios

Nome da Iniciativa de Origem: Metadata Requirements Project

Termo Adotado: Conseqüência do uso para efeitos probatórios

Definição: Identifica o impacto de determinado uso.

Observações: Deve identificar o impacto de certo tipo de uso. A mudança no conteúdo do arquivo ou em seus metadados pode significar a perda da autenticidade do mesmo.

ⁱ UNICODE: Padrão criado para permitir aos computadores representar e manipular texto de qualquer sistema de escrita existente.

ⁱⁱ *Bitstream* – um código de oito bits que serve para representar determinado dado dentro de um sistema computacional.