

Metadados para o gerenciamento eletrônico de documentos de caráter arquivístico - GED/A: estudo comparativo de modelos e formulação de uma proposta preliminar

Metadata for Electronic Records Management: a comparative study of models and formulation of a preliminary proposition

por [Katia P. Thomaz](#) e [Vilma Moreira dos Santos](#)

Resumo: O artigo apresenta uma análise comparativa dos modelos de metadados produzidos pelos principais projetos de manutenção de documentos eletrônicos de arquivo. Esses projetos foram utilizados como referencial teórico-metodológico para uma pesquisa de doutorado sobre a problemática do documento eletrônico de arquivo, em andamento na Escola de Ciência da Informação da Universidade Federal de Minas Gerais. Além de divulgar os estudos já realizados no escopo dessa pesquisa, o artigo tem como objetivo submeter à apreciação da comunidade acadêmica e dos profissionais ligados ao campo da arquivística um modelo preliminar de metadados para o gerenciamento eletrônico de documentos de caráter arquivístico. A expectativa é promover a discussão acadêmica em torno do tema e motivar o envio de críticas e sugestões que possam contribuir para o desenvolvimento da pesquisa.

Palavras-chave: Arquivologia; Gestão de Documentos de Arquivo; Gestão Eletrônica de Documentos de Arquivo; Documento Eletrônico de Arquivo; Metadados.

Abstract: This article presents a comparative study of metadata models proposed by some of the main projects on electronic recordkeeping. These projects were analysed in order to formulate the theoretical and methodological basis for a doctoral research project on the matter, which is being executed at School of Information Science of Federal University of Minas Gerais, Brazil. The aim of this article is to make public and discuss the preliminary results of the research.

Keywords: Archival Science; Records Management; Electronic Recordkeeping; Electronic Record; Metadata.

1. Introdução

Em janeiro de 1991, pesquisadores de várias áreas do conhecimento e de diferentes países se reuniram em Washington, por iniciativa da National Historical Publications and Records Commission - NHPRC, órgão vinculado ao National Archives and Records Administration - NARA. Os participantes desse encontro tiveram como objetivo debater temas relacionados ao gerenciamento de documentos eletrônicos de caráter arquivístico e produziram, ao final do encontro, uma agenda contendo vários itens a serem priorizados nas pesquisas da década que se iniciava (NARA/NHPRC, 2000).

O encontro de Washington pode ser considerado um marco no desenvolvimento da pesquisa no campo dos documentos eletrônicos de arquivo, pois, pela primeira vez, foi sistematizada uma agenda de pesquisa que se desdobrou em projetos em várias instituições acadêmicas e arquivísticas. Alguns desses projetos foram considerados fundamentais para compor o referencial teórico do projeto de pesquisa para admissão no PPGCI [1], nível de doutorado, levando-se em conta o caráter acadêmico dos mesmos e a sua repercussão em eventos da área.

Nesse artigo pretende-se desenvolver uma análise comparativa desses projetos, de modo a se obter um embasamento teórico-metodológico para a elaboração de um modelo conceitual de metadados, objetivo principal da pesquisa de doutorado, ora em andamento. Os projetos selecionados para estudo foram os seguintes:

* *Functional Requirements for Evidence in Recordkeeping* - 1993 a 1996 (coordenado por Richardo Cox, da Universidade de Pittsburgh, USA);

* *Strategic Partnerships with Industry - Research & Training – SPIRT* 1998 a 1999 (coordenado

por Sue McKemmish da Universidade de Monash, Austrália);

* ***International Research on Permanent Authentic Records in Electronic Systems - InterPARES*** - 1999 a 2001 (coordenado por Luciana Duranti, da Universidade de British Columbia, Canadá).

Tentando cobrir o escopo mínimo necessário para a compreensão dos assuntos tratados, o artigo foi estruturado nos seguintes tópicos:

- * o contexto do gerenciamento eletrônico de documentos de caráter arquivístico - GED/A
- * metadados para preservação de objetos digitais
- * o modelo de referência Open Archival Information System (OAIS), um ponto de partida potencial para a construção de uma estrutura de metadados de preservação digital
- * estudo comparativo dos projetos de manutenção de documentos eletrônicos de caráter arquivístico
- * um modelo preliminar de metadados para gerenciamento eletrônico de documentos de caráter arquivístico.

Cabe ressaltar o caráter preliminar desses estudos, uma vez que a pesquisa encontra-se em andamento. A nossa expectativa é promover a discussão acadêmica em torno do tema e motivar o envio de críticas e sugestões que possam contribuir para o desenvolvimento desta pesquisa [2].

2. O contexto do gerenciamento eletrônico de documentos de caráter arquivístico - GED/A

No ambiente manual, os documentos provenientes das funções ou dos processos de uma organização têm sido mantidos sempre juntos num "sistema de gerenciamento de documentos de arquivo". Tais sistemas têm funcionado como suporte aos responsáveis pelos processos e suas transações e têm, como principal objetivo, preservar os documentos em seus componentes individuais (isto é, conteúdo, estrutura e contexto) bem como sua relação dentro de uma série histórica de eventos para que se possam reconstruir fatos, eventos ou transações com a confiança necessária. **Essas características são cruciais para a preservação da proveniência e do valor de prova.** (NARA/NHRPC, 1999)

RECORDS MANAGEMENT		
TRADITIONAL RECORDS MANagements	ELECTRONICS RECORDS MANagements	
	Todos os suportes, inclusive papel	ELECTRONIC RECORDKEEPING Somente eletrônico
MANUAL	AUTOMATIZADO	

FIGURA 1 - O contexto do *records management* (gestão de documentos de arquivo), segundo o NARA/NHRPC (2000).

A FIG. 1 ilustra as disciplinas do campo do *records management*, num modelo proposto pelo

NARA/NHPRC (2000). Os termos dessa figura são assim conceituados:

* ***records management***: planejamento, controle, direcionamento, organização, treinamento, promoção e outras atividades gerenciais envolvidas na criação, manutenção, uso, e eliminação dos documentos de arquivo com vistas a atingir a adequada e correta documentação das políticas e transações do governo federal;

* ***electronic records management***: utilização de técnicas automatizadas para gerenciar documentos de arquivo, independentemente de seu formato; termo mais abrangente para se referir ao gerenciamento, por meios eletrônicos, de documentos em formatos variados, sejam eles eletrônicos, papel, microforma, etc;

* ***electronic recordkeeping***: processos automatizados para gerenciar os **documentos eletrônicos de arquivo** de uma agência do governo; tratam da preservação do **conteúdo, contexto e estrutura** do documento eletrônico **através do tempo**.

Segundo a proposta do NARA, o termo "records management" está abrigando a gestão de documentos nos moldes tradicionais, em que as operações se davam manualmente, como também o uso de recursos eletrônicos para o gerenciamento de documentos eletrônicos e outros suportes. Para esse último caso, o NARA propõe o termo "electronic records management", cujo conceito poderia ser representado em língua portuguesa pelo termo **gerenciamento eletrônico de documentos de caráter arquivístico e pela sigla GED/A**. Verifica-se, também, que o termo "electronic recordkeeping", ainda segundo o NARA, está sendo usado para designar a parte desse gerenciamento eletrônico que trata dos documentos efetivamente eletrônicos.

O estudo da literatura tem-nos levado a inferir que o termo "keeping" vem sendo empregado para se referir apenas à sistemática de manutenção de documentos de arquivo, excluindo os processos gerenciais de planejamento e administração. Os projetos analisados neste artigo adotam o termo "recordkeeping" (sem o adjetivo "electronic") para incorporar, dentro dessa sistemática de manutenção, os documentos em suporte não eletrônico.

Considerando o exposto, estamos adotando ao longo deste artigo as seguintes traduções:

* ***records management*** – gestão de documentos de arquivo;

* ***electronic records management*** – gestão eletrônica de documentos de arquivo ou simplesmente GED/A;

* ***recordkeeping*** – manutenção de documentos de arquivo;

* ***electronic recordkeeping*** - manutenção de documentos eletrônicos de arquivo.

Os projetos analisados neste artigo e os principais autores do campo da arquivística têm defendido a abordagem de metadados para o gerenciamento eletrônico de documentos de arquivo, por considerarem mais adequada à nova realidade do alto volume de informação e complexidade do ambiente eletrônico. O grande desafio está na identificação das características do mundo real e virtual (dos computadores) a serem preservadas.

3. Metadados para preservação digital

O termo metadados antecede a *Web* tendo, aparentemente, sido cunhado por Jack Myers nos anos 60 para descrever arquivos eletrônicos (MILSTEAD & FELDMAN, 1999) mas começou a aparecer mais

freqüentemente na literatura sobre Sistemas Gerenciadores de Banco de Dados (SGBD) nos anos 80, para descrever as características das informações armazenadas nos bancos de dados (VELLUCCI, 1998, p. 191).

No domínio dos SGBD's, os metadados são definidos simplesmente como "dados sobre dados". MILSTEAD & FIELDMAN (1999) defendem a idéia de que, na verdade, os metadados antecedem a existência do próprio termo e citam como um exemplo típico o conhecido formato MARC – Machine Readable Cataloging para descrever documentos. Nossa discussão, no entanto, se concentra nos sistemas que estão sendo desenvolvidos na atualidade, isto é, após o aparecimento das fontes eletrônicas.

O que são, afinal, metadados? O conceito está provando ser de difícil definição. A cada publicação, deparamos com uma definição diferente, direcionada a um tipo específico de aplicação ou uso. FERREIRA (1986) define o prefixo "meta" como primeiro elemento de compostos eruditos com a idéia de "mudança", "posterioridade", "além", "transcendência", "reflexão crítica sobre" o que nos remete à idéia de metalinguagem, ou seja, "linguagem para descrever outra linguagem ou qualquer sistema de significação: todo discurso acerca de uma língua, como as definições dos dicionários, as regras gramaticais, etc.". Esse conceito nos parece bastante apropriado para o ambiente da tecnologia da informação onde temos informação representada conforme regras e padrões definidos por uma linguagem de máquina e tudo o que precisamos são orientações para sua interpretação. Segundo nossa visão, os metadados constituiriam, portanto, tais orientações.

Wendler (segundo OCLC/RLG Working Group on Preservation Metadata, 2001, p. 2) observa que, no contexto dos objetos digitais, os metadados estão associados a uma de três categorias funcionais:

- * Descritivo: facilitar a descoberta e identificação de fonte;
- * Administrativo: suportar o gerenciamento de fonte dentro de uma coleção;
- * Estrutural: juntar os componentes de objetos de informação complexos.

Das três categorias, a que tem recebido maior atenção, notadamente através da iniciativa de metadados Dublin Core, é a categoria descritiva que, como vimos, visa solucionar, apenas, a pesquisa de fontes eletrônicas. Entretanto, conforme ressaltado pelo grupo OCLC/RLG, "a crescente preocupação dos desafios colocados pela preservação digital - a retenção de objetos digitais por longo prazo - tem ressaltado a necessidade de metadados além da descoberta de fonte".

Experiências anteriores no ambiente da computação indicam que o gerenciamento efetivo de objetos digitais tende a ser facilitado através da criação, manutenção e evolução de metadados. Os metadados podem, por exemplo, documentar procedimentos associados à preservação, especificar direitos de acesso e estabelecer a autenticidade do conteúdo digital. Podem unir os diversos componentes constituintes de um objeto complexo e, ainda, as diversas versões do mesmo objeto. Em resumo, **a criação e o desdobramento de metadados de preservação tende a ser um componente chave para a maioria das estratégias de preservação digital.**

Assim, uma estrutura de metadados de preservação digital deve descrever os tipos de informação que devem ser associados aos objetos digitais num ambiente de armazenamento e essa estrutura deve ser genérica, abrangente, estruturada e aplicável a uma vasta gama de objetos, atividades e instituições de preservação. Dentro dessa perspectiva o modelo de referência Open Archival Information System - OAIS tem-se destacado na comunidade internacional dos profissionais ligados a instituições culturais, já contando com várias iniciativas de implementação (OCLC/RLG Working Group on Preservation Metadata, 2001, p. 6).

4. O modelo de referência Open Archival Information System (OAIS)

A pedido da International Organization for Standardization (ISO), o Consultative Committee for Space Data Systems (CCSDS) - organismo internacional de cooperação entre agências de pesquisa interessadas no desenvolvimento de padrões para tratamento de dados espaciais - começou a coordenar uma iniciativa para o desenvolvimento de padrões arquivísticos para o armazenamento de dados em formato digital por longo-prazo. O CCSDS começou, então, a produzir um modelo de referência que permitiria estabelecer a terminologia e os conceitos básicos para descrever e comparar modelos de dados e arquiteturas de arquivos, identificar as entidades significativas e os relacionamentos entre entidades num ambiente de arquivos, esclarecer a função-chave e os componentes de informação de um sistema de arquivos e, finalmente, servir como um arcabouço (*framework*) para as atividades de construção de padrões.

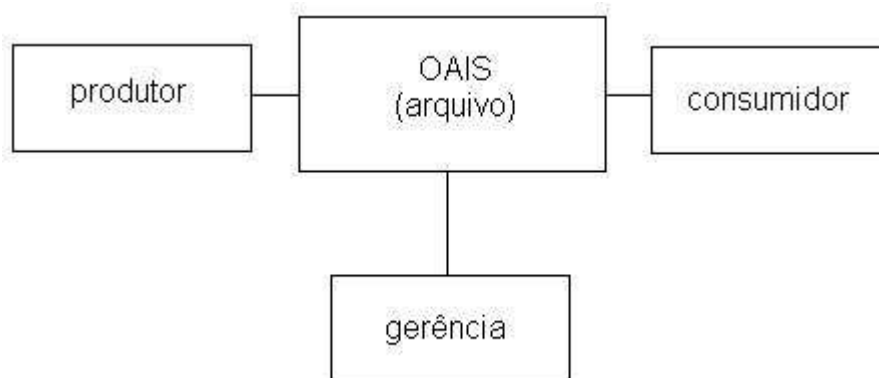


FIGURA 2 - O contexto de um OAIS (NASA/CCSDS, 2001)

Os trabalhos resultaram na publicação do modelo de referência Open Archival Information System (OAIS), em março de 1999. O modelo de referência é um arcabouço conceitual para um sistema de arquivos voltado para a preservação e a manutenção de acesso a informação digital por longo prazo. Descreve o ambiente, os componentes funcionais e a infra-estrutura de informação para suportar os processos do OAIS. O modelo passou por uma extensa revisão que envolveu, além da comunidade espacial, bibliotecas, arquivos e outras instituições culturais, agências de governo e o setor privado. Uma versão revisada foi publicada em julho de 2001 e a versão final em janeiro de 2002 [3].

O modelo de referência define um OAIS como "uma organização de pessoas e sistemas que aceitaram a responsabilidade de preservar informação e torná-la disponível a uma Comunidade Alvo" (NASA/CCSDS, 2001). Como pode ser observado pela FIG. 2, seu contexto é formado pela interação de três entidades: Produtores (*Producers*), Consumidores (*Consumers*), e Gerência (*Management*). Os Produtores fornecem as informações que o arquivo deve preservar. Os Consumidores são aqueles que utilizam tais informações. Uma categoria especial de Consumidores é a Comunidade Alvo - o subconjunto de consumidores que deve entender a informação preservada na forma armazenada. A Gerência é a entidade responsável pelo estabelecimento das políticas gerais do arquivo, dentro de um domínio mais abrangente de políticas.

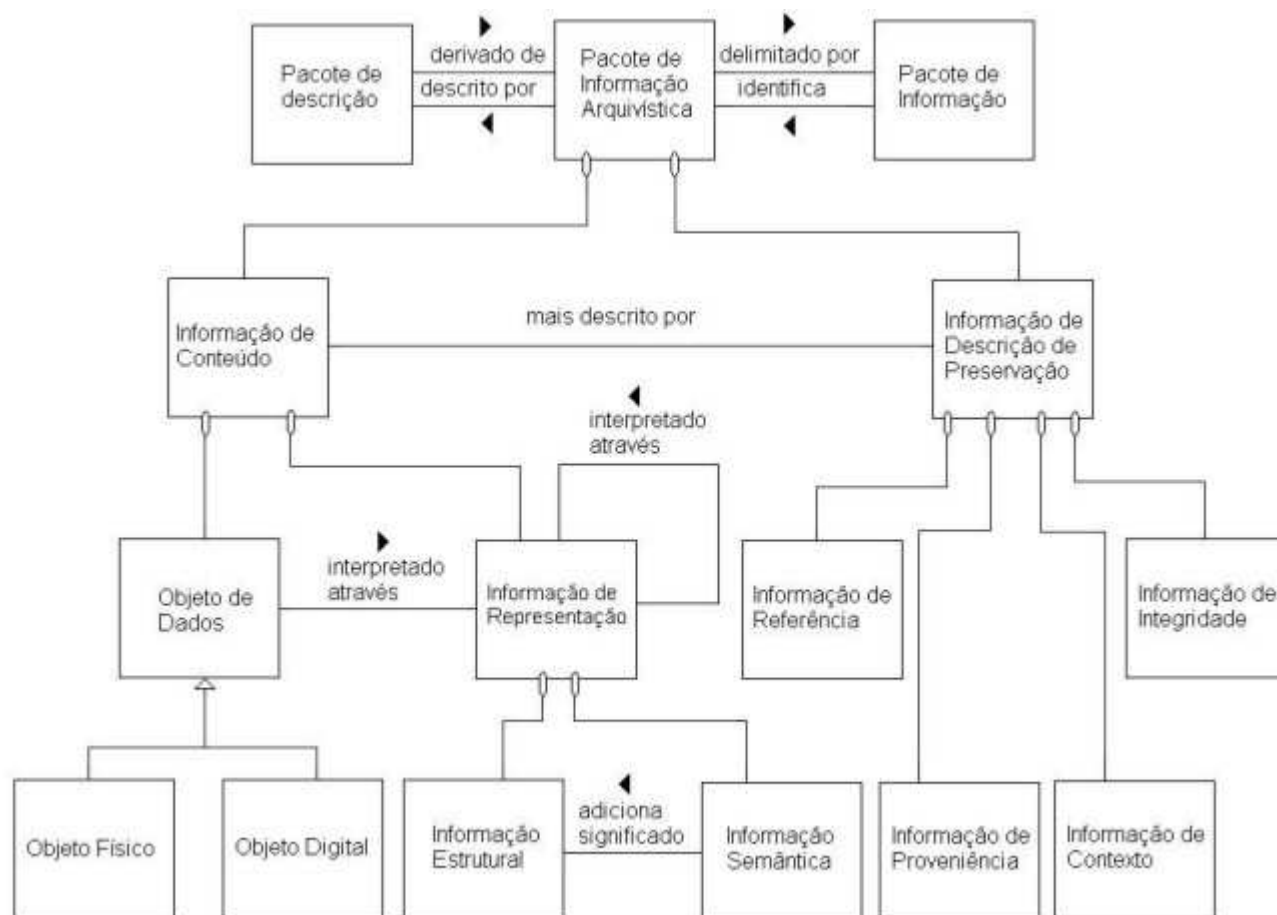


FIGURA 3 - O Pacote de Informação Arquivística - PIA (NASA/CCSDS, 2002)

A parte do modelo de referência que mais interessa para a questão dos metadados de preservação é o modelo de informação, cujo componente central é o Pacote de Informação Arquivística - PIA (*Archival Information Package - AIP*). O diagrama UML [4] da FIG. 3 ilustra o conceito.

O Pacote de Informação Arquivística é um continente [5] conceitual de dois tipos de informação: Informação de Conteúdo (*Content Information*) e Informação de Descrição de Preservação (*Preservation Description Information*). A Informação de Conteúdo e a Informação de Descrição de Preservação são unidas e identificadas pela Informação de Empacotamento (*Packaging Information*). A Informação de Conteúdo é a informação propriamente dita, alvo da preservação. Consiste do Objeto de Dados (*Data Object*) - os *bits*, no caso de objetos digitais - e sua Informação de Representação (*Representation Information*) associada, isto é, informação crítica responsável pela interpretação da cadeia de *bit's* do Objeto de Dados. A Informação de Representação pode ser Estrutural (*Structural*) ou Semântica (*Semantic*). A Informação de Descrição de Preservação, por sua vez, acrescenta significado à Informação de Conteúdo em quatro aspectos: Informação de Referência (*Reference Information*), Informação de Proveniência (*Provenance Information*), Informação de Contexto (*Context Information*) e Informação de Integridade (*Fixity Information*).

É importante destacar a natureza recursiva da Informação de Representação. Se a própria Informação de Representação estiver na forma digital, mais Informação de Representação será necessária para entender os seus *bit's* e assim sucessivamente até **um documento físico, de total compreensão humana, que dê início ao o processo de interpretação**. Destaca-se, ainda, que a interpretação de um Objeto de Dados depende não somente da Informação de Representação mas também da base de conhecimento dos Consumidores. Essa base de conhecimento, utilizada para entender e interpretar os dados, é externa ao modelo, não sendo, portanto, mantida, desenvolvida ou preservada como parte da função arquivística.

Dada a força do modelo de referência OAIS, em vias de tornar-se um padrão internacional (ISO 14.721), e sua aplicabilidade aos objetivos do projeto de pesquisa, decidiu-se adotá-lo como ponto de partida para a

comparação dos metadados dos modelos propostos pelos projetos analisados e elaboração de uma proposta preliminar.

5. Estudo comparativo dos projetos de gerenciamento de documentos de caráter arquivístico

Para desenvolvimento do estudo comparativo, os projetos foram analisados nos seguintes aspectos: justificativa e objetivos; premissas; metodologia de trabalho; estrutura de metadados; e elementos de metadados, agrupados segundo o modelo de referência Open Archives Information Systems - OAIS. Esse último aspecto não será apresentado neste artigo por se tratar de um conjunto muito extenso e intrincado de informações. Apenas a síntese dos elementos comparados, devidamente relacionados à sua origem por um padrão de cores, será descrita no modelo preliminar de metadados para gerenciamento eletrônico de documentos de caráter arquivístico.

5.1 Justificativas e objetivos

O projeto de Pittsburg estabeleceu como objetivo específico o desenvolvimento de um conjunto de requisitos funcionais [6] para sistemas de manutenção de documentos de arquivo [7] - atendendo, dentre outras, todas as necessidades legais e administrativas de uma organização específica. No decorrer do projeto buscou, também, identificar como a cultura e políticas organizacionais e a tecnologia da informação podem afetar as funções de manutenção.

O projeto SPIRT estabeleceu como objetivo específico a definição e a codificação de um conjunto genérico de metadados necessários para a manutenção de documentos de arquivo em ambientes distribuídos, ou seja, ambientes de rede local e/ou remota. Tais metadados devem garantir a manutenção da evidência e o acesso contínuo aos documentos eletrônicos de arquivo para fins de governabilidade (*governance*), prestação de contas (*accountability*), memória (*memory*) e identidade (*identity*). Ao término, o projeto recebeu um convite do *Standards Australia Committee IT/21* para participar da elaboração de um padrão australiano de metadados para manutenção de documentos de arquivo baseado no modelo produzido. Um subcomitê do IT/21, presidido por Adrian Cunningham, foi estabelecido para tratar do assunto.

O projeto InterPARES estabeleceu como objetivo geral o "desenvolvimento de conhecimento teórico e metodológico essencial à preservação permanente de documentos de arquivo gerados eletronicamente e, com base nesse conhecimento, formular um modelo de estratégias, políticas e padrões capazes de assegurar sua preservação". Este objetivo foi formulado ao se constatar que a capacidade de garantir a autenticidade, bem como a preservação por longo prazo, está ameaçada por fatores de natureza diversa, tais como: a natureza proprietária do *software* e meios de armazenamento, a obsolescência tecnológica, a grande diversidade de formatos, a grande complexidade dos relacionamentos físicos e lógicos no ambiente eletrônico, a fragilidade dos meios de armazenamento, a facilidade de manipulação dos sistemas eletrônicos e a coexistência de ambientes híbridos (convencional e eletrônico).

Em termos de objetivos é interessante verificar que os projetos podem ser considerados complementares. Enquanto o projeto de Pittsburgh se dedica aos estudos de requisitos funcionais, o projeto SPIRT busca a definição de metadados e o InterPARES volta-se para o desenvolvimento de políticas e estratégias. Na prática, entretanto, o projeto de Pittsburgh acabou modelando um conjunto de metadados que foi utilizado como base para o projeto SPIRT.

Em termos de foco, os projetos de Pittsburgh e SPIRT preocupam-se com os arquivos nas fases corrente e intermediária, por entenderem que uma solução nessas fases se refletirá naturalmente na fase subsequente - a permanente. O projeto InterPARES, por sua vez, preocupa-se com os arquivos na fase permanente, e defende a idéia de que, nesta fase, cessam os interesses de preservação por parte da organização produtora e, portanto, as instituições públicas devem assumir esse encargo em defesa dos interesses da sociedade. Nesse aspecto o projeto InterPARES se diverge radicalmente do projeto de Pittsburgh e, por conseguinte, do projeto SPIRT, uma vez que esses últimos acreditam na possibilidade de permanência dos documentos de arquivo nas organizações produtoras em sua fase permanente dentro da chamada "perspectiva de continuidade" (*continuum perspective*). (ERLANDSSON, 1996)

Dentro de uma visão da arquivística sistêmica integrada contemplada nos pressupostos da legislação arquivística brasileira, acredita-se que os **modelos devem ser adotados desde a fase corrente mas, de acordo com o projeto InterPARES, julgamos necessária a transferência definitiva dos documentos de arquivo para as instituições arquivísticas públicas em sua fase permanente.**

Na verdade, vamos um pouco além dessa idéia e, em função da natureza dinâmica e flexível dos negócios e do legado de *software* aplicativo em funcionamento dentro das próprias organizações produtoras, **recomenda-se a criação de um novo ambiente para a manutenção de documentos eletrônicos de caráter arquivístico considerados de valor permanente, ainda nas fases corrente e intermediária. Damos a esse novo ambiente o nome de Armazém de Documentos de Arquivo (*Records Warehouse*) [8], cuja função primordial é manter metadados para documentos de arquivo em formato convencional e eletrônico e, especificamente para os documentos eletrônicos, promover seu recolhimento e formatação conforme padrões de gerenciamento previamente definidos. Cabe ressaltar que esse recolhimento se dará no ambiente operacional (*hardware/software*) de geração desse documentos.**

5.2 Premissas e princípios fundamentais

O projeto de Pittsburgh parte das seguintes premissas:

- a) os requisitos funcionais para documentos de arquivo gerados eletronicamente são basicamente os mesmos considerados para documentos convencionais [9], mas vários requisitos não serão atendidos através dos métodos tradicionais desenvolvidos para o papel;
- b) as diferentes áreas de negócio apresentam diferentes graus de risco associados;
- c) os requisitos funcionais devem ser independentes de *software*;
- d) áreas de negócio similares terão os mesmos requisitos, mas a cultura corporativa determinará a escolha das táticas; e
- e) a cultura corporativa será fator determinante para o atendimento dos requisitos.

Um dos pontos-chave para entendimento da abordagem adotada pelo projeto SPIRT é a visão da gestão contínua dos documentos de arquivo (*continuum perspective*). Na gestão contínua, os documentos não são vistos como objetos passivos, descritos posteriormente no momento de sua transferência para arquivos permanentes. Os processos relacionados à descrição podem iniciar no momento ou, mesmo, antes da criação dos documentos e continuar ao longo de sua vida. No ambiente de rede e dos novos paradigmas emergentes da informação, o objeto informacional do tipo documento (*document-like information object* – DIO) pode, por si mesmo, tornar-se o agente da ação. Nessa condição de agente, ele pode provocar a tomada automática de decisões ou atos, previamente planejados por membros da organização e, ainda, registrar essa decisão ou ato. Dentro dessa ótica, as premissas fundamentais do projeto são as seguintes:

- a) é possível identificar, categorizar, rotular e apresentar de maneira formal e padronizada, metadados que suportem a manutenção de documentos de arquivo através do tempo e espaço – independentemente de onde, quando ou como esses metadados forem capturados; e
- b) é essencial desenvolver padrões de metadados para a manutenção de documentos de arquivo, compatíveis com as iniciativas desenvolvidas na comunidade mais abrangente de metadados.

O projeto InterPARES, por sua vez, formulou as seguintes premissas básicas:

- a) o projeto deve ser interdisciplinar e internacional, considerando o imenso escopo e abrangência das

questões em torno da preservação por longo prazo de autênticos documentos de arquivo;

b) a tecnologia, por si só, não é capaz de resolver todos os problemas ligados à preservação de documentos eletrônicos de arquivo por longo prazo (mesmo sabendo-se que os aspectos tecnológicos inerentes à questão tendem a ser explorados em inúmeros estudos e métodos de migração e que a indústria de *software* encontra-se, de forma crescente, preocupada em desenvolver arquiteturas de sistemas abertos e promover interoperabilidade entre tecnologias); e

c) na perspectiva da ciência arquivística, a despeito de novas mídias e formatos de documentos eletrônicos, os componentes essenciais que identificam e autenticam um documento de arquivo não se modificaram.

Nota-se que o projeto de Pittsburgh é o único a considerar em seus pressupostos a cultura corporativa como "fator determinante para o atendimento dos requisitos". O projeto SPIRT focaliza a questão dos metadados como aspecto central da preservação, bem como a interoperabilidade com outros padrões de metadados para futuras trocas de informação e utilização de padrões e ferramentas comuns. O projeto InterPARES destaca-se por retirar a ênfase na tecnologia voltando-se para o estudo e adaptação dos processos arquivísticos tradicionais.

Os três aspectos levantados pelos projetos são importantes para um ambiente eletrônico para manutenção de documentos de arquivo mas, a partir de nossos estudos, consideramos igualmente importantes os seguintes:

a) o documento eletrônico deve ser considerado dentro das seguintes perspectivas: documento originalmente eletrônico *versus* documento digitalizado para fins de preservação/aceso e documentos que registram os atos da organização (fatos ou eventos) *versus* documentos ou informações utilizadas exclusivamente para fins de processamento eletrônico (potencialidades);

b) a tecnologia deve servir ao homem, não o homem à tecnologia;

c) uma maior fragmentação no gerenciamento da informação arquivística deve ser evitada, buscando-se uma solução conjunta, que leve em conta os documentos convencionais e eletrônicos;

d) tendo em vista o caráter legal, cultural e histórico da informação arquivística, a dependência de *hardware* e *software* deve ser minimizada a partir do uso de padrões abertos;

e) tendo em vista as barreiras econômicas que assolam nossa região e os altos custos envolvidos na montagem de ambientes eletrônicos voltados para a manutenção segura de documentos de arquivo por longo prazo, as soluções conjuntas devem ser incentivadas.

5.3 Método de trabalho

O projeto de Pittsburgh definiu 13 (treze) requisitos funcionais de alto nível necessários para garantir a preservação de evidência em ambientes eletrônicos, extraídos de uma extensa revisão de literatura das áreas legal, auditoria, gestão de documentos de arquivo, computação e gerenciamento geral. Esses requisitos foram representados numa linguagem de regras de produção [10] passíveis de teste dentro de um sistema. Em seguida, procedeu à especificação dos metadados a serem associados e mantidos juntamente com o conteúdo e a estrutura do documento de arquivo, derivados dos requisitos funcionais e das regras de produção. A equipe do projeto foi composta por três grupos de participantes distintos: a) pessoas físicas e jurídicas que funcionavam como conselheiras do projeto; b) pessoas físicas e jurídicas que aplicavam e/ou testavam os produtos do projeto; e c) equipe do projeto, propriamente dita, composta por professores, estudantes e um consultor.

O projeto SPIRT desenvolveu um modelo conceitual de dados para os documentos de arquivo no seu contexto administrativo e sócio-legal em quatro etapas: a) análise da garantia literária, b) mapeamento dos

metadados; c) metamodelagem (entendido como modelagem de metadados); e d) teste com exemplos retirados do "mundo real". A análise da literatura envolveu a descoberta e estudo das fontes obrigatórias para a especificação de metadados de gestão de documentos de arquivo, padrões nacionais e internacionais, declarações de melhores práticas e os resultados das pesquisas. No mapeamento dos metadados, o projeto SPIRT produziu uma síntese dos elementos de metadados encontrados em conjuntos genéricos e específicos, destacando-se: o *Dublin Core - DC*, *Australian Government Locator Service - AGLS*, *Business Acceptable Communications Model - BAC*, *General International Standard Archival Description - ISAD (G)*, *International Standard Archival Authority Record for Corporate Bodies, Persons and Families - ISAAR*, *Encoded Archival Description Application Guidelines Version 1.0 - EAD*, conjunto de metadados do *Victorian Electronic Records Strategy - VERS* e *Recordkeeping Metadata Standard for Commonwealth Agencies* do National Archives of Australia. No que diz respeito à metamodelagem, foram empregadas duas técnicas formais de modelagem da ciência da computação, quais sejam, Resource Description Framework – RDF [11] e Object Role Modelling – ORM [12]. Conforme a documentação analisada, os objetivos da metamodelagem foram: permitir uma descrição rigorosa dos metadados; destacar as inconsistências e lacunas; proporcionar uma melhor especificação e relacionamento com outros modelos; e, finalmente, servir como um meio gráfico para divulgação. Os elementos de metadados foram descritos através dos atributos definidos na norma ISO/IEC 11179 [13] - Nome (*Name*), Identificador (*Identifier*), Versão (*Version*), Autoridade de Registro (*Registration Authority*), Idioma (*Language*), Definição (*Definition*), Obrigatoriedade (*Obligation*), Tipo de Dado (*Datatype*), Máximo de Ocorrências (*Maximum Occurrence*) e Comentários (*Comment*) - e atributos adicionais definidos no próprio projeto: Qualificador de Elemento (*Element Qualifier*), Componentes de Valor (*Value Components*), Esquema (*Scheme*), Descrição (*Description*) e Exemplos (*Examples*). O teste com exemplos do "mundo real" possibilitou destacar as inconsistências e lacunas da sintaxe e expressões semânticas dentro do conjunto e, também, fornecer orientações para potenciais implementadores para a aplicação da sintaxe.

Para investigar a complexa variedade das questões que afetam a preservação permanente de documentos autênticos no ambiente eletrônico, o projeto InterPARES definiu quatro campos interrelacionados de investigação, conduzidos por força tarefa específica: a) Requisitos Conceituais para Preservação de Autênticos Documentos Eletrônicos de Arquivo; b) Critérios e Metodologia para Avaliação de Autênticos Documentos Eletrônicos de Arquivo; c) Metodologias para Preservação de Autênticos Documentos Eletrônicos de Arquivo; e d) Estrutura para Desenvolvimento de Políticas, Estratégias e Padrões. Além dessas forças-tarefa, foi estabelecido um Comitê de Glossário para controlar o uso de vocabulário especializado dentro do projeto. As descobertas das forças-tarefa foram analisadas por uma equipe internacional que conduzia grande parte do trabalho através de um fórum de discussão na Web e que se reunia três vezes ao ano. Dependendo das questões de pesquisa, as forças-tarefa empregavam diferentes metodologias, tais como, análise diplomática, entrevistas estruturadas e análise de projeto de sistemas. Uma metodologia comum, no entanto, foi traçada para orientar todas as atividades de pesquisa: a modelagem. As forças-tarefa, com a ajuda de engenheiros de conhecimento [14], representavam suas descobertas em modelos para apresentação à equipe internacional. A metodologia de modelagem consiste, de maneira geral, de duas partes: representação gráfica das entidades [15] envolvidas, seus atributos [16] e relacionamentos, e identificação das atividades onde essas entidades estão envolvidas - a principal técnica de modelagem adotada no projeto foi o Integration Definition for Function Modelling - IDEF(0) [17]. A modelagem serve a vários propósitos mas principalmente para eliminar dúvidas na utilização de expressões, premissas, objetivos e interesses dos diferentes especialistas e estabelecer uma determinação rigorosa e sistemática do significado e implicação de cada termo, conceito ou declaração introduzida. Coube ao primeiro campo de investigação - Requisitos Conceituais para Preservação de Autênticos Documentos Eletrônicos de Arquivo - a tarefa de traduzir os requisitos em modelos para o desenvolvimento de metodologias e tecnologias a serem aplicadas em diversas instituições arquivísticas. Os resultados desses testes foram apresentados à equipe internacional para aperfeiçoamento dos modelos. Esse processo se deu de forma iterativa e teve como objetivo revelar alguns princípios básicos para a escolha das estratégias, políticas e padrões a serem traçados para a preservação de autênticos documentos eletrônicos de arquivo.

É importante notar que os três projetos constituíram-se de equipes multidisciplinares: profissionais dos arquivos e profissionais da computação (no caso do projeto InterPARES também profissionais da diplomática). Esse aspecto vem demonstrar a forte tendência de conscientizar as organizações para a

importância do trabalho conjunto nas atividades de concepção e projeto de sistemas eletrônicos de gerenciamento de documentos de arquivo. Esse caráter multidisciplinar trouxe, naturalmente, ferramentas e técnicas originalmente adotadas no âmbito da computação para a ciência arquivística: a modelagem de sistemas. O projeto de Pittsburgh utilizou técnicas da Inteligência Artificial (Regras de Produção), o projeto SPIRT acompanhou as tendências apontadas pelas iniciativas W3C (RDF e ORM) e o projeto InterPARES utilizou ferramentas derivadas da Análise e Projeto Estruturado de Sistemas (IDEF(0)).

Somos levados a crer que um modelo gráfico para representar as informações e as funções envolvidas na manutenção de documentos eletrônicos de arquivo é essencial para o entendimento e o compartilhamento de idéias entre técnicos e usuários. A força desse modelo estará na capacidade de representação dos diversos elementos envolvidos (funcionalidade), na facilidade de aprendizado/uso (usabilidade) e na abrangência de utilização. Considerando a proposta de criação de um Armazém de Documentos de Arquivo, mencionada no item 4.1, recomendamos a utilização das técnicas de modelagem de processos e de modelagem dimensional utilizadas no desenho de Armazém de Dados (Datawarehouse). É essencial, também, a descrição dos elementos de metadados segundo a norma ISO/IEC 11179 e sua implementação em XML [18] para garantir interoperabilidade entre os diferentes ambientes de *hardware* e *software*.

5.4 Estrutura de metadados

O modelo de metadados concebido pelo projeto de Pittsburgh, denominado Business Acceptable Communications - BAC, propõe seis camadas de metadados ligadas e retidas juntamente com o documento de arquivo. Essas camadas, conforme mencionado no item 4.3, foram projetadas para atender os requisitos funcionais de evidência em ambientes distribuídos de grande porte. São elas:

- a) Manuseio (*Handle Layer*): "declara" o documento de arquivo, indica seu contexto (em nível de transação) e fornece os termos para sua pesquisa;
- b) Termos e Condições (*Terms & Conditions Layer*): define os controles de acesso, uso e destinação (*disposition*);
- c) Estrutura (*Structural Layer*): detalha a estrutura de armazenamento e apresentação dos dados para permitir a manutenção de seu valor de prova e a migração para novo *software* e *hardware*;
- d) Contexto (*Contextual Layer*): identifica a origem (ou seja, a pessoa, sistema ou dispositivo responsável pela criação do documento de arquivo) e fornece dados de suporte para uso do documento de arquivo como evidência de uma transação;
- e) Conteúdo (*Content Layer*): contém os dados realmente envolvidos na transação;
- f) Histórico de Uso (*Use History Layer*): documenta os usos significativos do documento de arquivo após sua criação, incluindo descrição, reedições (*redacted releases*), e destinação/ eliminação (*disposition/ destruction*) autorizada pelo período de retenção.

O esquema de metadados definido no projeto SPIRT, *Australian Recordkeeping Metadata Schema – RKMS*, foi construído sobre um arranjo estrutural [19] que considera quatro classes de entidades primárias: Transações (*Business*), Agentes (*Agents*), Documentos-de-Arquivo (*Records*) e Processos-Arquivísticos (*Business-Recordkeeping*), bem como os Relacionamentos (*Relationships*) entre as entidades e os Instrumentos-Legais (*Mandates*) que governam as entidades e seus relacionamentos. O RKMS possibilita a captura de contexto de duas formas: como texto dentro de uma descrição de metadado, ou como um relacionamento com outras entidades. No segundo, deve-se tomar as devidas precauções para garantia da integridade dos relacionamentos. Qualquer dúvida a esse respeito pode indicar a implementação de metadados centrados nos documentos de arquivo [20]. A estrutura do RKMS prevê a descrição das entidades em diversos níveis de agregação. O esquema inclui uma lista completa dos níveis de agregação permitidos. Cada uma das entidades tem um conjunto próprio de metadados com dez elementos comuns: Categoria (*Category Type*), Identificador (*Identifier*), Nome (*Title*), Data (*Date*), Instrumento Legal

(*Mandate*), Local (*Place*), Classificação Funcional (*Functional Classification*), Relacionamento (*Relation*), Resumo (*Abstract*), Idioma (*Language*). Para a entidade Documentos-deArquivo foram definidos os seguintes elementos específicos: Classificação Temática (*Subject Classification*), Tipo Documental (*Documentary Form*), Avaliação (*Appraisal*), Controle (*Control*), Preservação (*Preservation*), Recuperação (*Retrieval*), Acesso (*Acess*), Uso (*Use*), Histórico (Event History). Os Relacionamentos e os Instrumentos-Legais ainda não foram descritos como entidades, permanecendo apenas como elementos. A sintaxe do RKMS suporta descrições de elementos, altamente estruturadas, com a possibilidade de desdobramento através do emprego de qualificadores [21], domínio de valor [22], referências a autoridades de registro [23] e sintaxe [24]. Prevê, ainda, a utilização de elementos de outros esquemas de metadados, permitindo que suas descrições possam ser estendidas através da estrutura e semântica de outros padrões de descrição.

O projeto InterPARES esboçou um modelo para análise (*template for analysis*) tendo como objetivo a identificação de todos os elementos conhecidos de um documento eletrônico de arquivo para posterior verificação de seu papel na confirmação da autenticidade. Os elementos foram agrupados num arcabouço (*framework*) que contempla: a) Tipo Documental (*Documentary Form*), b) Anotações (*Annotations*), c) Suporte (*Medium*) e d) Contexto (*Context*). O primeiro grupo - Tipo Documental - descreve as regras de representação do documento de forma que seu conteúdo, seu contexto administrativo e documentário e suas responsabilidades sejam comunicadas. Os elementos do Tipo Documental podem ser subdivididos em Elementos Extrínsecos, relativos a sua aparência externa e Elementos Intrínsecos, que armazenam o ato e seu contexto direto. O segundo grupo - Anotações - descreve os acréscimos feitos ao documento após sua criação. O terceiro grupo - Suporte - descreve o meio físico da mensagem. O quarto grupo - Contexto - descreve o ambiente jurídico-administrativo, organizacional e tecnológico, sob o qual acontece o ato.

Nota-se que as estruturas de metadados representam as opções de projeto para agrupamento dos diversos elementos e sofrem influências do referencial teórico adotado. A estrutura do projeto de Pittsburgh está bem segmentada e abrange todos os aspectos para descrição do documento eletrônico de arquivo, com exceção das características do suporte. A estrutura do projeto SPIRT é de difícil entendimento sem o estudo mais aprofundado de exemplos de sua utilização e parece bastante voltada para as informações de referência ao documento eletrônico de arquivo, talvez devido a sua origem no modelo AGLS. A estrutura do projeto InterPARES também está bem dividida e abrange todos os aspectos para descrição do documento eletrônico de arquivo, incluindo seu suporte.

5.5 Elementos de metadados

Visto que o modelo de referência OAIS pode ser considerado como ponto de partida para a construção de uma estrutura de metadados de preservação digital, utilizamos sua estrutura denominada de alto nível para fins de comparação dos modelos de metadados propostos pelos projetos analisados. Os elementos dos modelos *Business Acceptable Communications - BAC* do projeto de Pittsburgh, *Recordkeeping Metadata Schema - RKMS* do projeto SPIRT e do modelo para análise (*template for analysis*) do projeto InterPARES foram analisados e arranjados segundo a correspondência percebida com as categorias de informação do modelo OAIS. Com relação ao projeto InterPARES, faz-se necessário esclarecer que, no sentido estrito, os elementos do modelo para análise não foram identificados como elementos de metadados mas sim como "elementos conhecidos de um documento eletrônico de arquivo que podem influenciar na constatação de sua autenticidade".

Conforme mencionado no início do tópico 5, esta análise não será apresentada, tendo em vista a complexidade e a extensão de seu escopo. Entretanto, o resultado desta análise foi utilizado para desenvolvimento do tópico 6 a seguir. Cabe ressaltar que as interpretações são de inteira responsabilidade das autoras deste documento.

6. Modelo preliminar de metadados para o gerenciamento eletrônico de documentos de caráter arquivístico

Da síntese e crítica dos projetos analisados apresentamos, a seguir, uma proposta preliminar de metadados

para o gerenciamento eletrônico de documentos de caráter arquivístico. Cada elemento foi posicionado dentro do quadro, na linha correspondente à categoria de informação do modelo de referência OAIS - Estrutural, Semântica, Referência, Proveniência, Contexto, Integridade, Empacotamento -, conforme interpretação, e arranjado, hierarquicamente, numa estrutura de elementos e sub-elementos percebida visualmente através do recurso de identificação. Como mencionado anteriormente um padrão de cores (azul Pittsburgh; vermelho SPIRT, verde InterPARES, preto proposto pelas autoras) fará a ligação ao projeto de origem.

Categorias identificadas para um <i>Open Archival Information System</i> – OAIS	Elementos de Dados Sugeridos
1. Informação de Conteúdo: conjunto de informações que se deseja originalmente preservar	
1.1. Objeto de Dados: Objeto Físico ou Objeto Digital (cadeia de <i>bits</i>) que se deseja preservar	Conforme a literatura arquivística tradicional, o elemento atômico de um sistema de arquivos é o Documento de Arquivo e assim deve permanecer para o ambiente eletrônico, a despeito de possíveis detalhes técnicos de implementação. Caso se faça necessário, devem ser negociadas estratégias de formatação no momento da aquisição de material pelo sistema de manutenção de documentos eletrônicos de arquivo.
1.2. Informação de Representação: facilita a formatação, entendimento e interpretação de um Objeto de Dados	
1.2.1. Informação Estrutural: descreve a estrutura do Objeto de Dados; no caso de um objeto digital, interpreta os <i>bits</i> através da organização dos mesmos em tipos de dados específicos, grupos de tipos de dados e outros significados de alto nível (por exemplo, formato dos dados, ambiente de <i>hardware</i> e <i>software</i> necessário para acessar os dados)	<p>1.2.1.1. Identificação do Arquivo de Dados [25] (<i>File Identification</i>): possibilita a identificação individual dos arquivos de dados que compõem o documento e permite verificar sua autenticidade.</p> <p>1.2.1.1.1. Identificador do Arquivo de Dados: identifica cada arquivo de dados que compõe o documento; permite que o sistema reúna todas as partes do documento para formar o todo.</p> <p>1.2.1.1.2. Codificação do Arquivo de Dados (<i>File Encoding</i>): descreve a codificação de cada arquivo de dados contido no documento.</p> <p>1.2.1.1.2.1. Modalidade (<i>File-Modality</i>): identifica a modalidade do arquivo de dados (texto, numérico, gráfico, geográfico, imagem, som, vídeo, multimídia, etc).</p> <p>1.2.1.1.2.2. Representação dos Dados (<i>File-Data-Representation</i>): identifica os padrões de codificação dos dados utilizados (ASCII , EBCDIC , ou UNICODE character data , ASN.1 , CCITT Group III raster, etc).</p> <p>1.2.1.1.2.3. Código dos Dados (<i>Data-Codes</i>): indica, para os casos onde não foram usados métodos de representação padronizados, especificamente como os dados estão codificados; por exemplo, para dados vetor se <i>topological</i>, <i>spaghetti</i>, <i>chain-node</i>, etc., para dados raster a quantidade de pontos por polegada e densidade de bit, para dados por amostragem a quantidade de amostras por segundo, etc.</p>

1.2.1.2.4. Método de Compressão (*Compression-Method*): identifica, quando for o caso, o método de compressão utilizado (ex. Nenhum, JPEG, MPEG, Quiktime, LZW, etc.); se o método atende a um padrão específico, deve conter somente a identificação deste padrão (nome, versão, etc.), caso contrário, pode ser necessário definir tecnicamente o método.

1.2.1.2.5. Método de Criptografia (*Encryption-Method*): identifica os algoritmos empregados para criptografar o conteúdo do documento; todos os documentos de são armazenados na forma "descriptografada", como teriam sido vistos pelos receptores.

1.2.1.3. Apresentação do Arquivo de Dados (*File Rendering*): identifica aspectos específicos para que o documento possa ser devidamente visualizado no momento do recebimento.

1.2.1.3.1. Dependência de Aplicação (*Application Dependency*): indica, quando for o caso, as aplicações das quais o documento depende; se existirem dependências, o nome da aplicação, versão e informação de registro são documentados em cada ocorrência do campo no momento da criação do documento; essa informação deve servir como ponteiro para uma biblioteca registrada mantida pela organização produtora ou uma entidade pública tal como o Registro de Patentes.

1.2.1.3.2. Dependência de Software (*Software-Environment-Dependency*): indica, quando for o caso, o *software* básico, incluindo sistemas operacionais e API's, do qual o documento é dependente; se existir uma dependência, o nome do(s) pacote(s) de *software*, a versão, informação de registro e informação de apresentação (tais como grupos de fontes ou outros atributos dependentes de software) são documentados no momento da criação do documento.

1.2.1.3.3. Dependência de Hardware (*Hardware-Dependency*): indica, quando for o caso, o *hardware* do qual o documento é dependente; se existir uma dependência, o *hardware* necessário, modelo, número, configuração e informação de saída (tais como, impressoras ou visualizadores necessários ou outros atributos de dependência de *hardware*) são documentados no momento da criação do documento.

1.2.1.3.4. Regras de Apresentação (*Renderring-Rules*): identifica os procedimentos necessários para que o documento seja apresentado na tela, impressora ou de outra forma representado como no momento da criação (macros, dimension, dados geo-referenciados, etc.) – pode operar em níveis diferentes.

1.2.1.3.5. Padrão de Representação (*Representaton-Standard/De Facto Standard*): identifica padrão(ões),

	<p>incluindo a versão, aplicado(s) ao arquivo de dados que afetam sua forma de apresentação (ex: SGML, Postscript, TIFF, etc).</p> <p>1.2.1.4. Apresentação do Documento de Arquivo (<i>Record Rendering</i>): aplica-se ao documento como um todo, uma vez que os arquivos de dados tenham sido apresentados corretamente de acordo com sua regra própria.</p> <p>1.2.1.4.1. Regra ou Padrão de Ligação dos Arquivos (<i>File-Linking-Rule/Standard</i>): identifica as regras ou padrões necessários para possibilitar as ligações necessárias entre os arquivos de dados que produzem o documento; contem informação textual a respeito das regras ou padrões aplicados.</p> <p>1.2.1.4.2. Padrão para Intercâmbio de Dados: Versão (<i>File-Interchange-Standard: Version</i>): identifica o(s) padrão(ões), incluindo a versão, empregado(s) pelo documento para possibilitar o intercâmbio de arquivo de dados.</p>
<p>1.2.2. Informação Semântica: acrescenta significado às estruturas de dados identificadas pela Informação Estrutural (por exemplo, idioma)</p>	<p>1.2.2.1. Linguagem (<i>Language</i>): Nome do idioma ou <i>script</i> do documento</p> <p>1.2.2.2. Regras ou convenções: indica as regras ou convenções internacionais, nacionais e/ou locais seguidas na preparação do documento de arquivo (sinais, símbolos, abreviaturas, glossário de termos)</p>
<p>2. Informação de Descrição de Preservação: acrescenta informação que, embora não seja crítica para o entendimento da Informação de Conteúdo, permitirá, por um período indefinido de tempo, o entendimento do ambiente no qual a Informação de Conteúdo foi criada.</p>	
<p>2.1. Informação de Referência: enumera e descreve identificadores associados à Informação de Conteúdo de forma que possa ser referenciada de forma única, interna e externamente</p>	<p>2.1.1. Identificador (<i>Identifier</i>): identificador único associado ao documento dentro de um domínio específico (organização, município, estado, região, país, internacional).</p> <p>2.1.2. Tipo documental: tipo de documento em termos de fórmula diplomática, natureza de conteúdo ou técnica do registro, tais como cartas precatórias, cartas régias, cartas patentes, decretos-leis, decretos sem número, decretos legislativos, fotografias, litogravuras, serigrafias, xilogravuras.</p> <p>2.1.3. Identificação original/cópia/eletrônico</p> <p>2.1.4. Número: número ou código associado ao tipo documental</p> <p>2.1.5. Título (<i>Title</i>): nome do documento</p> <p>2.1.6. Data (<i>Date</i>): especifica hora/data/intervalo de data associado ao documento e "eventos da vida" relacionados à mesma</p> <p>2.1.7. Local (<i>Place</i>): informação sobre o local associado ao documento</p>

	<p>2.1.8. Pessoa(s): pessoa(s) associada(s) ao documento</p> <p>2.1.8.1. Papel: tipo de associação da pessoa ao documento (autor, relator, revisor, aprovador, subscritor, testemunha, emissor, destinatário, receptor)</p> <p>2.1.8.2. Nome: nome da pessoa associada ao documento</p> <p>2.1.8.3. Titulação(ões): título(s) apresentado(s) para a pessoa no documento</p> <p>2.1.2.9. Classificação funcional (<i>Functional Classification</i>): informação sobre a classificação funcional da entidade de acordo com uma fonte autorizada, por exemplo, uma Instituição ou Esquema de Classificação Funcional ou Tesauro.</p> <p>2.1.2.10. Classificação(ões) temática(s) (<i>Subject Classification</i>): informação sobre a classificação ou indexação do assunto referente ao conteúdo dos documentos de acordo com uma fonte autorizada, por exemplo, um Esquema de Classificação de Assunto, Lista de Cabeçalhos para Indexação ou Tesauro.</p> <p>2.1.2.11. Resumo (<i>Abstract</i>): descrição textual livre (não estruturada) do documento.</p>
<p>2.2. Informação de Proveniência: documenta o histórico da Informação de Conteúdo (por exemplo, sua origem, cadeia de custódia, atos e efeitos da preservação</p>	<p>2.2.1. Organização produtora: organização produtora do documento de acordo com uma fonte autorizada, por exemplo, uma Instituição ou Esquema de Classificação de Entidades</p> <p>2.2.2. Fonte (<i>Source</i>): identifica a fonte do documento e documenta as circunstâncias relevantes da captura dos dados.</p> <p>2.2.2.1. Fonte de Dados (<i>Data-Source</i>): identifica a fonte que produziu o documento; por exemplo o sistema de gerenciamento de documentos de arquivo.</p> <p>2.2.2.2. Documentação do Sistema Fonte de Dados (<i>Data-Source-System-Documentation</i>): identifica ou consiste da documentação que descreve as condições necessárias para produzir o documento – contem informações sobre o processamento dos dados.</p> <p>2.2.2.3. Tipo de Dispositivo de Captura de Dados (<i>Data Capture-Instrument-Type</i>): identifica o tipo de dispositivo que foi utilizado para a captura dos dados (<i>light recording, sound recording, temperature recording, location recording, etc</i>) e o dispositivo específico utilizado (fabricante, modelo, etc.).</p> <p>2.2.2.4. Calibragem do Dispositivo de Captura de Dados (<i>Data Capture-Instrument-Settings</i>): identifica os ajustes, calibragem, etc. no momento da captura dos dados.</p>

	<p>2.2.2.5. Qualidade da Fonte de Dados (<i>Source Data-Quality</i>): identifica o grau de confiabilidade dos dados gerados pela fonte.</p> <p>2.2.3. Histórico (<i>Use History</i>): descreve os eventos ou atos relevantes passados, presentes ou futuros relacionados ao documento.</p> <p>2.2.3.1. Tipo (<i>Use-Type</i>): identifica como os dados foram usados: visualizados, copiados, editados, armazenados, indexados, classificados, enviados, descartados, etc.; os diversos tipos de uso permitidos pelo sistema devem ser identificados.</p> <p>2.2.3.2. Data (<i>Use-Instance-Time</i>): identifica quando os dados foram usados – ou seja, a data e hora em que foram usados.</p> <p>2.2.3.3. Responsável (<i>Use-Instance-User</i>): identifica quem ou o que usou os dados numa determinada data/hora.</p> <p>2.2.3.4. Parecer (<i>Use-Evidential Consequences</i>): identifica o impacto de um uso específico (por exemplo, pode identificar a parte do documento liberada, os termos usados na indexação, a importância de uma visão específica, que parte do registro foi vista).</p>
<p>2.3. Informação de Contexto: documenta os relacionamentos da Informação de Conteúdo com seu ambiente (por exemplo, porque foi criado, relacionamentos com outras Informações de conteúdo)</p>	<p>2.3.1. Contexto de Proveniência (<i>Provenancial Context</i>): descreve o órgão de criação, seu mandato, estrutura e funções.</p> <p>2.3.1. Contexto Jurídico-administrativo (<i>Juridical-Administrative Context</i>): descreve o sistema legal e organizacional ao qual a organização produtora pertence.</p> <p>2.3.2. Contexto Transacional (<i>Procedural Context</i>): descreve o procedimento organizacional onde o documento foi criado.</p> <p>2.3.3. Contexto Documentário (<i>Documentary Context</i>): descreve os fundos ao qual o documento pertence e sua estrutura interna.</p> <p>2.3.5. Contexto Tecnológico de Criação: descreve as características dos componentes técnicos do sistema eletrônico onde o documento foi criado.</p> <p>2.3.6. Contexto Tecnológico de Armazenamento (<i>Technological Context</i>): descreve as características dos componentes técnicos do sistema eletrônico onde o documento encontra-se armazenado.</p>
<p>2.4. Informação de Integridade: documenta mecanismos de autenticação usados para garantir que a Informação</p>	<p>2.4.1. Criação: indica os procedimentos aplicáveis ao processo de criação do documento de arquivo.</p> <p>2.4.2. Avaliação: indica os procedimentos aplicáveis ao processo de avaliação do documento de arquivo.</p>

<p>de Conteúdo não foi alterada de forma não documentada (por exemplo, soma de fechamento, assinatura digital [26])</p>	<p>2.4.3. Acesso: indica os procedimentos aplicáveis ao acesso do documento de arquivo.</p> <p>2.4.4. Uso: indica os procedimentos aplicáveis ao uso do documento de arquivo.</p> <p>2.4.5. Eliminação: indica os procedimentos aplicáveis ao processo de eliminação do documento de arquivo.</p> <p>2.4.6. Transferência: indica os procedimentos aplicáveis ao processo de transferência do documento de arquivo.</p> <p>2.4.7. Validação (<i>Attestation</i>): validação escrita de um documento por aqueles que tomaram parte na sua publicação (autor, relator, contra-assinante) e pelas testemunhas do ato ou da assinatura do documento.</p> <p>2.4.8. Corroboração (<i>Corroboration</i>): menção explícita aos meios utilizados para validar o documento.</p> <p>2.4.9. Garantia do Sistema: indica os procedimentos aplicados para garantia do funcionamento adequado do sistema.</p> <p>2.4.9.1. Tipo: tipo de procedimentos (log, auditoria, bloqueio de acesso, etc).</p> <p>2.4.9.2. Data: data/hora/intervalo de data associado ao procedimento</p> <p>2.4.9.3. Responsável: pessoa ou entidade responsável pelo procedimento</p> <p>2.4.9.4. Parecer: relato dos fatos encontrados e recomendações cabíveis</p>
<p>3. Informação de Empacotamento: une ou relaciona, realmente ou logicamente, a Informação de Conteúdo e a Informação de Descrição de Preservação, numa entidade identificável, numa mídia específica</p>	<p>3.1. Unidade de arquivamento: descreve o tipo de suporte, endereçamento lógico do documento de arquivo dentro do sistema de arquivo e indica os procedimentos para a embalagem e conservação do suporte.</p> <p>3.1.1. Tipo: tipo de unidade física de arquivamento do documento (caixa, pasta, disquete, CD, etc)</p> <p>3.1.2. Endereço Lógico/Físico: possibilita a identificação individual da unidade de arquivamento dentro do sistema de arquivos</p> <p>3.1.3. Acondicionamento: indica os procedimentos aplicáveis ao acondicionamento da unidade de arquivamento</p> <p>3.1.4. Conservação: indica os procedimentos de rotina necessários para a conservação da unidade de arquivamento</p>

Tendo em mente uma aplicação em nível mundial, seria interessante associar ao identificador do documento de arquivo uma identificação única para o sistema eletrônico de gerenciamento de documentos de caráter arquivístico ou, conforme nossa proposição, para o Armazém de Documentos de Arquivo (*Records Warehouse*). O nome do sistema poderia ser formatado de acordo com um padrão que oferece aos usuários ou outros arquivos parceiros a informação necessária para estabelecer uma conexão, como o padrão ISO X.500 Directory Services Naming. (NASA/CCSDS, 2001)

Há, também, um aspecto de suma importância a ser destacado no que diz respeito ao item 1.2 (Informação de Representação). O modelo de referência OAIS afirma que "a preservação de informação por prazo indeterminado exige uma descrição completa e compreensível da Informação de Representação". Exemplificando com um documento produzido em processador de texto recomenda que "se o formato do processador de texto for proprietário e não puder ser adquirido, mesmo no nível de um simples visualizador de documento, é necessário migrar o documento para um formato não proprietário para assegurar sua Preservação por Longo Prazo" (NASA/CCSDS, 2001). Sobre essa questão a comunidade arquivística deve acompanhar o desenvolvimento do padrão Portable Document Format/Archive - PDF/A cujo "principal objetivo é oferecer 'padrões para preparação de documentos e metadados associados para fins arquivísticos', ou seja, permitir que as páginas de um documento armazenadas como um PDF/A sejam vistas da mesma forma daqui a 50 anos". Se tudo correr bem, o padrão deve ser publicado pela ISO em dois anos. Os primeiros documentos estão programados para 18 de novembro de 2003. (DUHON, 2002)

Levando-se em conta as possibilidades de um projeto individual de pesquisa, com o nível de aprofundamento desejável, a etapa seguinte desta pesquisa buscará detalhar e validar, para o caso brasileiro, as condições necessárias ao adequado monitoramento das tecnologias (*hardware/software/mídia*) envolvidas no funcionamento de um Armazém de Documentos de Arquivo (*Records Warehouse*). Acredita-se que aos elementos factuais inseridos no item 1.2.1.3 (Apresentação do arquivo) devam ser associados indicadores gerenciais que apontem, com um grau confiabilidade desejável, a necessidade de planejamento de ações preventivas de preservação (migração, conversão, emulação, dentre outras).

Notas

[1] THOMAZ, Katia P. *A guarda de documentos eletrônicos de arquivo*; novos desafios, velhos problemas. Belo Horizonte, 2002. (Projeto de pesquisa)

[2] Interessados poderão enviar críticas/sugestões através do endereço de correio eletrônico katia.thomaz@uol.com.br.

[3] Para maiores detalhes a respeito do modelo consultar NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA). Consultative Committee for Space Data Systems (CCSDS). *Reference model for an Open Archival Information System (OAIS); draft recommendation*. Washington, 2001.

[4] **UML - Unified Modeling Language**, é uma notação padrão para modelagem de objetos do mundo real, como primeiro passo no desenvolvimento de uma metodologia de análise e desenho de sistemas orientada-a-objeto.

[5] Julgamos mais adequada a tradução do termo inglês "container" para o português "continente", que segundo o Dicionário Aurélio Eletrônico Século XXI - Versão 3 - Novembro de 1999, significa "aquilo que contém algo" (3), sentido exato que desejamos expressar.

[6] Para o projeto de Pittsburgh, **requisitos funcionais** (*functional requirements*) são o conjunto de elementos necessários para garantir o adequado funcionamento de um sistema de manutenção de documentos de arquivo ou documentos eletrônicos de arquivo, extraídos da literatura (Arquivística, Computação, Direito e Administração), de padrões nacionais e internacionais, da legislação e das necessidades específicas de uma organização.

[7] Para o projeto de Pittsburgh, **sistemas de manutenção de documentos de arquivo** (*recordkeeping systems*), são "sistemas que criam, identificam, capturam, mantêm e usam documentos de arquivo, independente se na forma eletrônica ou papel". (UNIVERSITY OF PITTSBURGH. School of Information Sciences, 2001)

[8] A escolha do nome se deu por uma analogia ao termo Armazém de Dados (*Data Warehouse*) da computação que trata-se de um repositório central para todos ou parte dos dados significativos colecionados pelos diversos sistemas de negócios de uma organização. Um Armazém de Dados, tipicamente, reside num servidor de grande porte (mainframe) de uma organização. Os dados das diversas aplicações de processamento de transação on-line (*on-line transaction processing - OLTP*) e outras fontes são seletivamente extraídos e organizados no banco de dados do armazém para uso por aplicações analíticas e consultas de usuários (<www.whatis.com> em 22 jan. 2003).

[9] Documentos convencionais de arquivo são aqueles em formatos variados não eletrônicos, como papel, microforma, etc. (NARA/NHPRC, 2000)

[10] A equipe do projeto de Pittsburgh utilizou uma técnica do campo da inteligência artificial para expressar formalmente cada requisito funcional numa linguagem de **regras de produção**. Espera-se que essas regras possam ser testadas em avaliações de sistemas de manutenção de documentos de arquivo. (UNIVERSITY OF PITTSBURGH. School of Information Sciences, 2001)

[11] *Resource Description Framework* é "uma iniciativa do *World Wide Web Consortium* para permitir a criação, intercâmbio e uso de metadados. O RDF fornece um arcabouço (*framework*) sobre o qual as comunidades independentes podem desenvolver vocabulários de metadados que atendam suas necessidades específicas e compartilhar esses vocabulários com outras comunidades. O RDF define uma linguagem para descrição desses vocabulários, influenciada pelas idéias de representação de conhecimento da inteligência artificial e das comunidades de banco de dados". (W3C RDF Homepage: <<http://www.w3.org/RDF>>)

[12] **Object-Role Modelling** é "um método para o desenho de modelos de base de dados. É uma abordagem de modelagem conceitual, ou seja, especifica o modelo utilizando conceitos e linguagem de fácil entendimento para usuários leigos. Vislumbra o mundo em termos de objetos e papéis que desempenham. Através de uma técnica específica de desenho de esquema conceitual, esses objetos e papéis são identificados e expressos em sentenças de linguagem natural elementar. O ORM é particularmente adequado à modelagem de um esquema de metadados. Em primeiro lugar, porque é mais expressiva do que outras (tal como a modelagem de entidade-relacionamento) e essa expressividade permite um nível mais elevado de detalhamento e, conseqüentemente, uma análise mais rigorosa. Em segundo lugar, porque um diagrama ORM pode ser povoado com amostras de dados, permitindo a validação por um especialista utilizando linguagem natural e exemplos reais." (UNIVERSITY OF PITTSBURGH. School of Information Sciences, 2001)

[13] **ISO/IEC 11179-1** *Information technology - Specification and standardization of data elements - Part 1-6*. "Para facilitar as comunicações eletrônicas globais, a comunidade de Padrões Internacionais tem insistentemente trabalhado para definir um *Open Systems Interconnection Environment* (OSIE) dentro do qual diferentes hardware e aplicações possam compartilhar informações. Padrões vêm sendo propostos ou definidos para três (*hardware, software* e comunicações) dos quatro (*hardware, software, comunicações* e dados) componentes básicos necessários para sistemas abertos de processamento de informações. A norma ISO/IEC 11179 para especificação de dados, constitui o quarto componente para sistemas abertos de informação, oferecendo mecanismos para permitir que dados sejam compartilhados no OSIE". (ISO/IEC 11179, 1999)

[14] **Engenheiro do conhecimento** é o "cientista da informática que constrói um sistema especialista através da aquisição do conhecimento necessário e da conversão desse conhecimento em um programa" (MICROSOFT PRESS, 1998). As características exigidas para esse profissional geralmente são obscuras sendo, a maioria deles, portanto, auto-didatas. PAYNE & AWAD (citados por BARACSKAI &

VALENCEI, 2002) esclarecem que seus conhecimentos e habilidades incluem: conhecimento da tecnologia da computação, técnicas gerais de localização de fatos, métodos de prototipação, fatores humanos, áreas funcionais, habilidades de comunicação, habilidades de planejamento de projeto, habilidades de relacionamento humano, habilidades organizacionais e qualidades pessoais.

[15] Uma **entidade** é "qualquer coisa de interesse, concreta ou abstrata, incluindo associações entre coisas". (ISO/IEC 11179, 1999).

[16] Um **atributo** é uma característica de um objeto ou entidade. (ISO/IEC 11179, 1999)

[17] **IDEF(0)** - *Integration Definition for Function Modelling* é um US Federal Information Processing Standard para elaboração de modelos funcionais, detalhado na Publicação 183 do *National Institute of Standards of Technology* - NIST de 21 de dezembro de 1993. Segundo este padrão, "um modelo funcional é uma representação estruturada das funções, atividades ou processos dentro do sistema modelado ou área de assunto".

[18] **eXtensible Markup Language** - XML é um formato de texto bastante flexível derivada da SGML (ISO 8879), originalmente desenhada para atender os desafios da publicação eletrônica em grande escala (W3C, 2003). XML vem desempenhando, de forma crescente, um importante papel no intercâmbio de uma grande variedade de dados na Web e qualquer outro ambiente. Segundo ROCKLEY (2002), suas principais vantagens são: (1) possibilita a determinação de uma estrutura previamente definida, (2) tem como meta principal "tornar documentos transportáveis através de sistemas e aplicações", (3) oferece liberdade para definição de nomes de rótulos (tags), (4) foi especificamente projetada para trabalhar com bases de dados; (5) apresenta um poderoso mecanismo, XSL (eXtensible Style Language), tanto para transformar quanto para formatar documentos XML; e (6) permite adaptações e personalizações.

[19] Um **arranjo estrutural** é um "método de posicionamento de objetos em um contexto demonstrando seus relacionamentos com outros objetos". Exemplos incluem modelos de entidade-relacionamento, taxonomias e ontologias. (ISO/IEC 11179, 1999)

[20] **Metadados centrados nos documentos de arquivo** significa que todas as informações de contexto serão transportadas ou reproduzidas dentro dos limites do sistema de manutenção de documentos de arquivo propriamente dito, evitando-se qualquer referência a fontes ou sistemas externos. (MONASH UNIVERSITY. School of Information Management and Systems. Records Continuum Research Group, 2000)

[21] **Qualificador** é "um termo que ajuda a definir e construir um conceito único" (ISO/IEC 11179, 1999), como por exemplo, nascimento para data de nascimento.

[22] **Domínio de valor** é "o conjunto de valores permitidos" para um elemento de metadado. (ISO/IEC 11179, 1999)

[23] **Autoridade de Registro** é "uma organização autorizada a registrar elementos de dados ou outros objetos" (ISO/IEC 11179, 1999), onde deverão, futuramente, ser depositados os Esquemas de Metadados formais (poderíamos, talvez, fazer uma analogia com a Autoridade de Registro da Infra-Estrutura de Chaves Públicas do Governo Brasileiro e a assinatura digital).

[24] **Sintaxe** é "uma estrutura de expressões de uma linguagem, e as regras que governam essas estruturas; relacionamentos entre caracteres ou grupo de caracteres, independentemente do seus significados ou da forma de interpretação e uso". (ISO/IEC 11179, 1999)

[25] Julgamos mais adequada a tradução do termo inglês "file", no sentido da computação, para o português "arquivo de dados". Um arquivo de dados é, portanto, uma entidade discreta, disponível aos usuários do sistema (incluindo o próprio sistema operacional e seus programas aplicativos). Para ser identificado, um

arquivo de dados precisa ter um nome único dentro de um domínio (diretório) específico. Alguns sistemas operacionais e aplicações descrevem arquivos de dados de determinados formatos através de um sufixo específico (também conhecido como extensão do arquivo).

[26]A abordagem mais básica [de **assinatura eletrônica**] é simplesmente escrever seu nome. [...] Outro tipo de abordagem é capturar e apresentar uma imagem digitalizada da assinatura pessoal dentro do documento para ilustrar a aprovação do conteúdo do documento e identificar o assinante. [...] Algumas aplicações também coletam dados biométricos sobre a pressão e o traço da assinatura e, então, fecham o documento de forma que nenhuma mudança possa anular a assinatura e as características da assinatura possam ser associadas ao usuário.

A maioria das soluções de assinatura exigem software aplicativo tanto no computador do assinante quanto do receptor. Isso pode funcionar bem dentro de uma organização mas poderia não ser prático entre indivíduos ou organizações não vinculados. Na tentativa de preencher essa grande necessidade surgem as **assinaturas digitais** [grifo nosso] utilizadas dentro de uma infra-estrutura de chave pública ou ICP [...] que repousam no complexo mundo da criptografia assimétrica." (MINIHAM, 2002)

Referências Bibliográficas

ARQUIVO NACIONAL. Conselho Nacional De Arquivos. *Resolução n. 4 - 28 mar. 1996. Dispõe sobre o Código de Classificação de Documentos de Arquivo para a Administração Pública: Atividades-Meio, a ser adotado como modelo para os arquivos correntes dos órgãos e entidades integrantes do Sistema Nacional de Arquivos (SINAR), e aprova os prazos de guarda e a destinação de documentos estabelecidos na Tabela Básica de Temporalidade e Destinação de Documentos de Arquivo Relativos às Atividades-Meio da Administração Pública.*

BAILEY, C. *Archival theory and machine readable records; some problems and issues*. Vancouver: University of British Columbia, 1988. (dissertação, mestrado).

BARACSKAI, Zoltán, VELENCEI, Jolán. *Important characteristics for a knowledge engineer*. Budapest: Doctus Bt. (artigo capturado na Internet em 14 dez. 2002 no endereço <http://www.doctus.info/white/3.doc>).

DUHON, Bryant. PDF/A. *e-doc*, v. 16, n. 6, nov./dez. 2002. (capturado na Internet em 4 jan. 2003 no endereço <http://www.edocmagazine.com/>).

ERLANDSSON, Alf. *Electronic records management; a literature review*. Paris: International Council Of Archives (ICA), 1996.

FERREIRA, Aurélio B. H. *Novo dicionário da língua portuguesa*. 2.ed. rev. aum. Rio de Janeiro: Nova Fronteira, 1986.

ISO/IEC FDIS 11179-1: 1999, *Information technology – Specification and standardization of data elements – Part 1: Framework for the specification and standardization of data elements*.

MICROSOFT PRESS. *Dicionário de informática*. Rio de Janeiro: Campus, 1998.

MILSTEAD, Jessica; FELDMAN, Susan. Metadata: project and standards. *Online*, v. 23, p. 32-40, jan./feb.1999.

MINIHAM, Jim. Digital signatures. *e-doc*, v. 16, n. 13, May/June 2002, p. 34-35.

MONASH UNIVERSITY. School of Information Management and Systems. Records Continuum Research Group. *Recordkeeping Metadata Project*. Last updated on June 23, 2000. (capturado na Internet em 27 nov. 2001 no endereço <http://rcrg.dstc.edu.au/research/spirt/index.html>).

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA). Consultative Committee for Space Data Systems (CCSDS). *Reference model for an Open Archival Information System (OAIS)*; draft recommendation. Washington, 2001.

NATIONAL ARCHIVES AND RECORDS ADMINISTRATION (NARA). National Historical Publications and Records Commission (NHPRC). *Context for electronic records management (ERM)*. Last Modified on April 11, 2000 (página acessada na Internet em 24 ago. 2000 no endereço: <http://www.nara.gov/records/fasttrak/prod1afn.html>).

ROCKLEY, Ann. Managing Enterprise Content. *e-doc*, v. 16, n.4, July/Aug. 2002, p. 32-35.

ROUSSEAU, J.Y.; COUTURE, C. *Os fundamentos da disciplina arquivística*. Lisboa: Dom Quixote, 1998. 356p.

UNIVERSITY OF PITTSBURGH. School of Information Sciences. *Functional Requirements for Evidence in Recordkeeping*. (capturado na Internet em 27 nov. 2001 no endereço <http://www.sis.pitt.edu/~nhprc>).

UNIVERSITY OF BRITISH COLUMBIA. School of Library, Archival and Information Studies. *InterPARES Project*. (capturado na Internet em 13 jan. 2003 no endereço <http://www.interpares.org>).

WORLD WIDE WEB CONSORTIUM (W3C). *eXtensible Markup Language (XML)*. (capturado na Internet em 28 jan. 2003 no endereço <http://www.w3.org/XML/>).

Sobre as autoras / About the Authors:

Katia P. Thomaz
katia.thomaz@uol.com.br

Bacharel em Ciência da Computação pela UFMG
Especialização em Engenharia da Qualidade pela PUCMG
Certified Document Imaging Architect - CDIA™ pela CompTIA
Certified Project Management Professional - PMP™ pelo PMI
Aluna do Programa de Pós-Graduação em Ciência da Informação - nível Doutorado, da Escola de Ciência da Informação da Universidade Federal de Minas Gerais.

Vilma Moreira dos Santos
vmsantos@eci.ufmg.br
Doutora pela University of Liverpool / Inglaterra
Professora do Programa de Pós-Graduação em Ciência da Informação da UFMG e orientadora do projeto em pauta.