

# Documentos eletrônicos de caráter arquivístico: fatores condicionantes da preservação

Katia P. Thomaz

Doutora em Ciência da Informação pela UFMG e-mail:  
katia.thomaz@uol.com.br

A pesquisa sobre a preservação de documentos eletrônicos de caráter arquivístico, desenvolvida no Programa de Pós-Graduação em Ciência da Informação da ECI-UFMG, através de abordagem transdisciplinar, investigou o ambiente de arquivamento de informação digital em organizações públicas brasileiras de médio e grande portes para identificar aqueles fatores condicionantes de sua preservação. Este artigo apresenta visão geral da pesquisa, esquematiza os objetivos, metodologia, resume a análise conceitual e discute os principais resultados.

Palavras-chave: Arquivologia; Gestão de documentos; Documentos eletrônicos; Preservação digital

Recebido em 28.04.2005

Aceito em 13.05.2005

“Mas uma vez percorrido, este caminho permanece traçado e lhe parece evidente, e esqueces o milagre de uma caminhada que foi semelhante a um retorno.”  
(Saint-Exupéry, *Cidadela*)

## Introdução

A crescente proliferação de documentos digitais, sobretudo nato-digitais<sup>1</sup>, tem ameaçado a capacidade humana de continuar utilizando os arquivos como fontes confiáveis de informação em virtude dos novos desafios impostos pela preservação. Nessa conjuntura de incerteza é possível perguntar: “Que fatores condicionam<sup>2</sup> esse novo ambiente no sentido de assegurar que o documento digital cumpra o seu percurso natural de vida, da administração à história?” Se estamos perdendo informação em formato digital, o que estaria ocasionando essa perda? Seria a falta de preparo do consumidor dessa informação ou o descuido com a manutenção dos recursos tecnológicos? Ou a carência de políticas de preservação na valorização e na priorização de medidas preventivas e conseqüente disponibilidade de recursos financeiros?

Essas reflexões e suas conseqüências em longo prazo para a história e a identidade da nação constituíram inspiração básica para o desenvolvimento da tese *A preservação de documentos eletrônicos de caráter arquivístico: novos desafios, velhos problemas*<sup>3</sup>.

## Aspectos metodológicos

A inexistência de metodologia de pesquisa consolidada no campo ensejou o delineamento de processo específico. Inicialmente, dois pressupostos básicos foram estabelecidos. Em primeiro lugar, devido a complexidade do ambiente no qual o documento eletrônico se insere, foi preciso elaborar o modelo sintético do ambiente da preservação digital detalhado na FIG. 1, baseado em teorias organizacionais (BLAU & SCOTT, 1970; MIRANDA, 1978; HALL, 1984; CHAMPION, 1985; CHIAVENATO, 1987; BOWDITCH e BUONO, 1997), para identificar os fatores condicionantes na fase de coleta de dados.

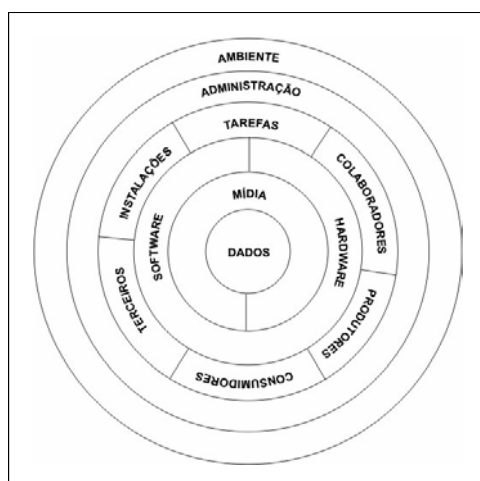


FIGURA 1 - Ambiente da preservação digital - modelo simétrico

1 Documento nato-digital: documento criado originalmente no ambiente informático cujo manuseio efetivo só pode, em alguns casos, ser feito no mesmo ambiente.

2 Fator condicionante é definido na tese como “certa imposição ou restrição, resultante de circunstâncias ou de decisão prévia, que deve ser observada na solução de um problema”.

3 Defendida em dez. 2004 no Programa de Pós-graduação em Ciência da Informação da UFMG, linha de pesquisa Informação, Cultura e Sociedade.

Esse modelo, embora apresentado como pressuposto, não deixa de ser também um dos resultados da pesquisa. O modelo contempla 13 classes de objetos<sup>4</sup>, a saber:

- ⊙ *Dados arquivados*: dados produzidos e mantidos para acesso futuro;
- ⊙ *Mídia de armazenamento de dados arquivados*: os diversos tipos de materiais físicos nos quais os dados foram gravados e armazenados, como disquetes, discos rígidos, fitas e discos ópticos;
- ⊙ *Software de apresentação de dados arquivados*: *software* que apresenta todos os dados arquivados ou parte dos mesmos, em formatos compreensíveis por pessoas ou sistemas-cliente;
- ⊙ *Hardware para processamento de dados arquivados*: *hardware* necessário para o processamento da mídia de armazenamento de dados e do *software* de apresentação de dados;
- ⊙ *Tarefas de manutenção*: conjunto de atividades necessárias para a conservação dos dados, da mídia de armazenamento de dados, do *software* de apresentação de dados, do *hardware* para processamento de dados e da instalação;
- ⊙ *Tarefas de negócio*: conjunto de atividades (meio e fim) desenvolvidas na organização em decorrência de seu negócio;
- ⊙ *Colaboradores*: conjunto de pessoas responsáveis pela execução das tarefas de manutenção;
- ⊙ *Produtores de dados*: conjunto de pessoas ou sistemas-cliente que produzem os dados a serem arquivados;
- ⊙ *Consumidores de dados arquivados*: conjunto de pessoas ou sistemas-cliente que têm interesse nos dados arquivados;
- ⊙ *Terceiros*: conjunto de organizações ou pessoas que fabricam produtos, fornecem serviços ou certificam dados;
- ⊙ *Instalações*: conjunto de aparelhos ou peças que compõem determinada utilidade;
- ⊙ *Administração*: estrutura organizacional e demais elementos de trabalho necessários para o desempenho das tarefas de negócio; e
- ⊙ *Ambiente*: conjunto de fatores e condições externas que têm relevância imediata para a organização.

As relações de dependência entre as classes devem ser consideradas a partir do setor mais interno da figura em direção à sua periferia. De fato, os dados dependem da mídia onde estão registrados, que por sua vez dependem do *software* para serem interpretados e do *hardware* para serem lidos e processados, que por sua vez dependem da interação entre tarefas, colaboradores, produtores, consumidores, terceiros e instalações para gerenciá-los, que por sua vez dependem da administração para estabelecer políticas de preservação, que finalmente é influenciada pelo ambiente geral.

<sup>4</sup> *Classe de objetos*: conjunto de coisas, reais ou abstratas, reconhecidas como sendo do mesmo tipo por compartilharem as mesmas características de atributos, operações, relações e semântica (FURLAN, 1998).

Em segundo lugar, pela dificuldade de estudar casos reais de preservação digital por longo prazo<sup>5</sup>, foi preciso estabelecer a mudança tecnológica como objeto inicial de análise. Previu-se que o estudo dos diversos elementos que seriam encontrados nas mudanças tecnológicas<sup>6</sup> do ambiente informático das organizações revelaria um conjunto de fatores vitais relacionados à preservação digital. Por quê? Porque essas mudanças constituem momentos de grande risco para a preservação dos dados, pois os mesmos devem passar por procedimento de migração do ambiente no qual se encontram controlados e estáveis, para novo ambiente, sem que ele produza a perda de suas características originais.

Os objetivos básicos da investigação visaram ampliar o conhecimento no campo da preservação de documentos eletrônicos de caráter arquivístico, de forma a favorecer sua necessária transferência a futuras gerações. Pôde-se constatar que não existem no Brasil estudos mais aprofundados sobre o tema. Esse objetivo geral, se atingido, poderia propiciar o surgimento de novos trabalhos e pesquisas que complementaríamos a iniciativa. De forma específica, buscou-se identificar fatores a serem monitorados em ambiente de arquivamento de documentos eletrônicos para garantir sua preservação por longo prazo. Ou seja, onde concretamente se deve atuar para reduzir os riscos de perda dos registros digitais?

A definição do escopo e dos objetivos básicos conduziram a pesquisa à abordagem qualitativa que se caracteriza sobretudo pela ausência de medidas numéricas e análises estatísticas, examinando-se aspectos mais profundos e subjetivos do tema. O pesquisador age como intérprete da realidade, desenvolvendo conceitos, idéias e entendimentos a partir de padrões encontrados nos dados ao invés de coletá-los para comprovar teorias, hipóteses e modelos preconcebidos. Optou-se pela combinação de estudo de caso, *grounded theory*, entrevista, análise documental e grupo focal, sendo que uma abordagem foi complementando a outra.

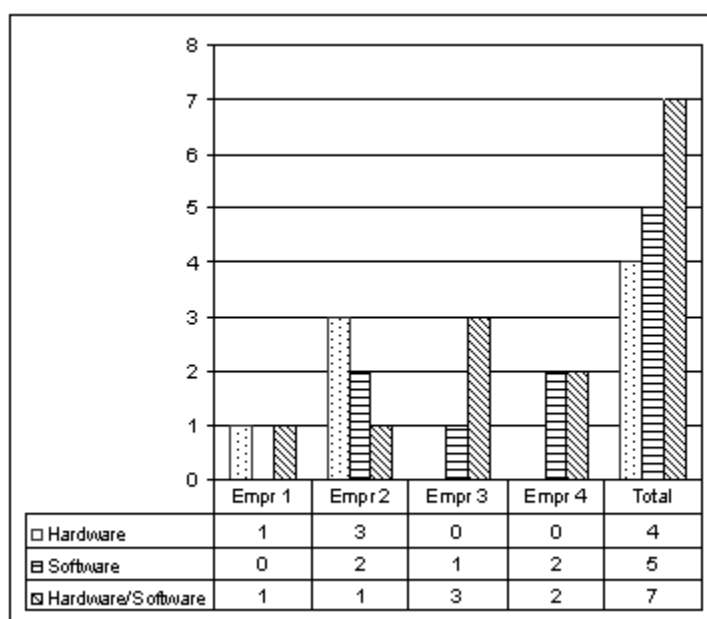


GRÁFICO 1- Quantidade de mudanças tecnológicas investigadas por tipo de componente.

<sup>5</sup> Constatou-se através de contatos acadêmicos e profissionais durante a pesquisa: inexistência de experiências práticas brasileiras de preservação digital na perspectiva da abordagem pretendida, observado através de consultas às organizações para fins de sondagem; pouca substância prática nos trabalhos produzidos na América do Norte, Europa e Austrália; caráter inovador do tema da pesquisa nos cursos brasileiros de graduação e pós-graduação das escolas da área; e estruturação incipiente dos órgãos normativos brasileiros dos arquivos com relação ao documento eletrônico, demonstrada por Santos (2002).

<sup>6</sup> Processo no qual componente(s) de *hardware/software* é (são) substituído(s) em nível de marca, modelo, versão, *release* e atualização, que poderá ser considerado como *obsolescência provocada*.

Participaram da pesquisa as empresas: Companhia Energética de Minas Gerais - CEMIG, Companhia de Saneamento Básico de Minas Gerais - COPASA, Empresa de Informática e Informação do Município de Belo Horizonte - PRODABEL e Companhia de Tecnologia da Informação do Estado de Minas Gerais - PRODEMGE. Elas detêm grande parte do patrimônio de informações arquivísticas digitais de interesse do Estado de Minas Gerais e do Município de Belo Horizonte, na medida em que são os principais responsáveis quanto à guarda e à administração dessas informações. Outro aspecto relevante na escolha foi o potencial interesse dos resultados do estudo por outras empresas de energia, saneamento básico e informática dos estados brasileiros, devido à similaridade de seus modelos de informação. Assim, do universo de mudanças tecnológicas promovidas com sucesso pelas empresas pesquisadas no período 2000-2002<sup>7</sup>, dezesseis foram selecionadas para análise dos registros técnicos produzidos em variados suportes convencionais e digitais<sup>8</sup> durante o processo de mudança – 71 documentos, 522 páginas –, que constituíram os primeiros objetos para a coleta de dados. O GRAF. 1 mostra a distribuição das mudanças investigadas por empresa e tipo de componente envolvido na mudança.

<sup>7</sup> Três anos foi considerado o prazo ideal para a pesquisa, em decorrência das previsões de ciclos médios de renovação de tecnologia apontados na literatura. Hedstrom (1997/1998) e Brand (1999) apontam ciclos de três a cinco anos.

<sup>8</sup> Esses registros envolveram: pequenos documentos como atas, memorandos, ofícios, e documentos de grande extensão como editais e relatórios técnicos.

<sup>9</sup> Cargo de média gerência ligado à indústria de informática: gerencia, planeja e coordena as atividades das áreas de suporte aos usuários; incluindo o desenvolvimento de estudos de *hardware* e *software*; controla o desempenho dos sistemas e recursos técnicos instalados visando a correção de possíveis irregularidades; propõe estudos sobre modificações, aperfeiçoamento, desenvolvimento e ajustes no sistema operacional dos equipamentos para atenderem às novas exigências dos usuários.

<sup>10</sup> Cargo técnico ligado à indústria de informática: planeja e coleta informações junto aos usuários para implantar sistemas de processamentos de dados; desenvolve sistemas a partir da análise de informações, estudando fluxos de trabalho e necessidades de recursos para propor alterações de rotina e elaborar propostas; implanta e mantém sistemas observando eficiência, racionalidade e solução de problemas técnicos; elabora manuais de sistemas.

<sup>11</sup> Cargo técnico ligado à indústria de informática: participa na análise, estudo, seleção, planejamento, instalação, implantação e manutenção de *software* básico e de apoio, como sistemas operacionais, banco de dados, teleprocessamento etc.; participa no estudo, implantação e documentação de rotinas para a melhoria das operações do computador; padroniza a análise e a programação dos sistemas de aplicação; acompanha o desempenho dos recursos técnicos instalados.

Duas categorias de sujeitos foram identificadas nas empresas com funções diretamente ligadas à questão da pesquisa. A primeira categoria envolveu quatro gerentes de suporte técnico<sup>9</sup> ou cargo similar da organização, entrevistados na fase inicial de coleta de dados. A segunda categoria abrangeu 40 especialistas da computação – analistas de sistemas<sup>10</sup> e analistas de suporte técnico<sup>11</sup>, com mais de cinco anos de experiência na função e com vivência em, no mínimo, uma mudança tecnológica –, oriundos do quadro de colaboradores das empresas pesquisadas, que se dispuseram a participar voluntariamente das avaliações dos fatores condicionantes da preservação digital.

Através do processo interativo de coleta, análise, crítica, consolidação e ordenação dos dados – ciclo repetido na busca de maior refinamento e qualidade das informações obtidas –, a pesquisa evoluiu a partir dos dados brutos encontrados nos registros de mudanças tecnológicas até à última interação com os especialistas da computação para avaliar a relevância de um conjunto de fatores condicionantes da preservação digital.

## Discussão dos resultados

O processo de investigação espelhou as relações informacionais encontradas no campo da pesquisa através de contínuo diálogo entre pressupostos e fatos, e culminou com a identificação de 160 fatores condicionantes da preservação digital (ANEXO A). Foi elaborado um glossário detalhando o significado das classes, fatores e termos relacionados. O GRAF. 2 mostra a distribuição dos fatores segundo as classes do modelo sintético do ambiente da preservação digital, definido como primeiro pressuposto da pesquisa (FIG. 1). De fato, os números revelaram que, em ordem decrescente de quantidade de fatores, as classes Ambiente, Tarefas de manutenção, Dados e Mídia deverão exigir os maiores esforços de preservação.

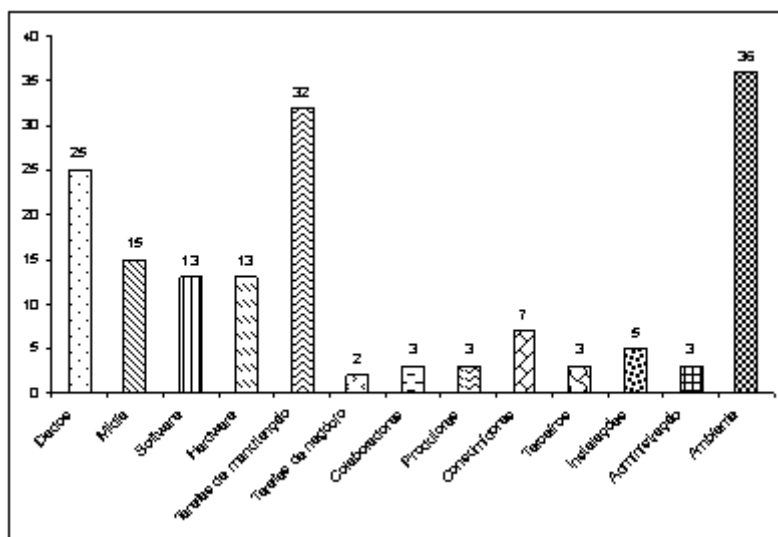


GRÁFICO 2- Distribuição dos fatores condicionantes da preservação digital por classe de objetos

Buscando a validação e a integração dos fatores condicionantes da preservação digital no campo mais abrangente da pesquisa, esses fatores foram examinados à luz de três aspectos fundamentais da literatura de preservação digital. Inicialmente, os fatores foram comparados com os problemas identificados por diversos autores (CONWAY, 1996; WATTERS & GARRET, 1996; INTERNATIONAL COUNCIL ON ARCHIVES - ICA, COMMITTEE ON ELECTRONIC RECORDS - CER, 1997; HEDSTROM, 1997/1998; BRAND, 1999; BULLOCK, 1999; PUBLIC RECORDS OFFICE - PRO, 1999; ROTHENBERG, 1999; BEAGRIE & JONES, 2002; EUROPEAN COMMISSION - EC, INTERCHANGE OF DATA BETWEEN ADMINISTRATIONS - IDA, 2001; THIBODEAU, 2001; LUSENET, 2002; NATIONAL LIBRARY OF AUSTRALIA - NLA, 2003)<sup>12</sup>, apurando-se a distribuição apresentada no GRAF. 3. Os resultados indicaram que as maiores concentrações se encontram, em ordem decrescente, na vulnerabilidade física com 61 fatores, na vulnerabilidade lógica com 49 fatores e na obsolescência tecnológica com 27 fatores. As vulnerabilidades física e lógica constituem, portanto, as maiores preocupações e não a obsolescência tecnológica, como indicavam as primeiras previsões.

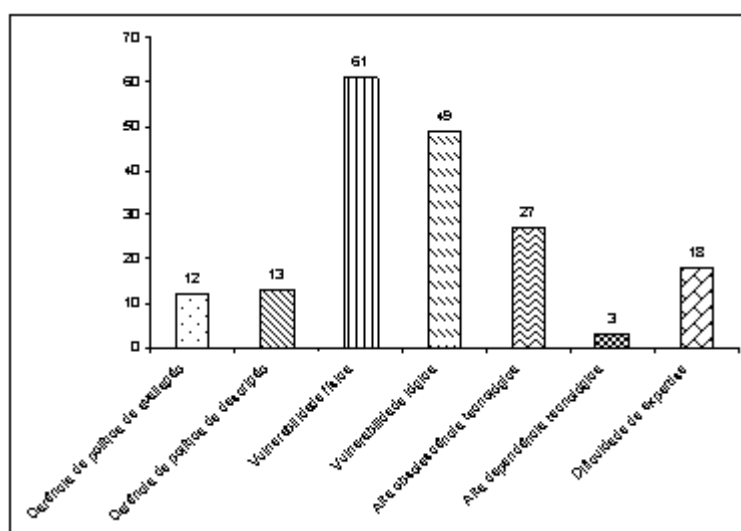


GRÁFICO 3- Distribuição dos fatores condicionantes da preservação digital por tipo de problema

<sup>12</sup> Para fins de verificação e classificação dos fatores condicionantes, os problemas fundamentais da preservação digital foram reunidos em sete categorias, a partir do ponto de vista da organização:

I *Carência de políticas de avaliação*: As implicações para definição de prioridades e critérios de seleção são muito graves no ambiente informático. A fonte digital não selecionada para tratamento ativo de preservação nas fases iniciais tem grandes possibilidades de ser perdida ou se tornar inutilizável no futuro;

II *Carência de políticas de descrição*: A natureza complexa da tecnologia exige que seja adotada abordagem de descrição mais detalhada dos objetos digitais para a sua manutenção. O envolvimento, tanto interno quanto externo, torna-se essencial;

III *Vulnerabilidade física*: O hardware e as mídias nas quais os objetos digitais são armazenados são inerentemente instáveis e, sem ambientação e manutenção adequadas, podem deteriorar muito rapidamente, mesmo que não pareçam danificados externamente;

IV *Vulnerabilidade lógica*: O ambiente informático é propício à mudança – algumas mudanças decorrem de necessidades do próprio gerenciamento do ambiente – o que pode comprometer a integridade, a autenticidade e a história dos objetos digitais;

V *Alta obsolescência tecnológica*: A velocidade das mudanças tecnológicas no ambiente informático, reduz o tempo para a necessária reação. O ciclo de renovação é curto – de três a cinco anos –, ao contrário de décadas ou mesmo séculos associados à preservação de objetos físicos. A obsolescência tecnológica é geralmente vista como a principal ameaça técnica para garantir o acesso continuado ao objeto digital;

VI *Alta dependência tecnológica*: Todos os objetos digitais requerem hardware e software específicos para serem acessados. Essa tecnologia – ampla e, na maioria das vezes proprietária – requer ajustes contratuais adequados nem sempre de fácil negociação; e

VII *Dificuldade de obtenção de expertise*: A tecnologia envolvida no acesso a objetos digitais requer grande diversidade de mão-de-obra especializada e atualizada para seu funcionamento e sua manutenção.

<sup>13</sup> Baseando no modelo de referência *Open Archival Information System - OAIS*, Bullock (1999) identificou nove ações, chamadas de requisitos, a serem observadas, de forma progressiva na preservação de documentos eletrônicos:

I *Fixar os limites do objeto a ser preservado*: Embora a natureza multimídia e hipertextual dos objetos digitais seja bastante vantajosa do ponto de vista da navegação, para fins de preservação é necessário definir, claramente, que elementos serão efetivamente mantidos;

II *Preservar a presença física*: A presença física representa o(s) arquivo(s) físico(s), isto é, a camada primitiva de suporte da informação a ser representada; refere-se, portanto, ao(s) arquivo(s) de computador, às séries de Os e Is que são a base para o significado do objeto digital;

III *Preservar o conteúdo*: Refere-se a manter a capacidade de acessar o conteúdo em seu nível mais baixo, como um arquivo texto em ASCII, independentemente do estabelecimento de variações de fontes e características de leiaute;

IV *Preservar a apresentação*: O conteúdo é apresentado visualmente através da aplicação de fontes de diferentes formatos e tamanhos, uso de espaço em branco, colunas, margens, cabeçalhos, rodapés, paginação e assim por diante. Em alguns tipos de documentos digitais, como formatos padrão SGML e alguns formatos PDF, as especificações de apresentação ficam separadas do conteúdo;

V *Preservar a funcionalidade*: Objetos digitais podem conter componentes multimídia, isto é, conter texto, gráficos, áudio e vídeo integrados; existir em formato hipertexto, isto é, com capacidade de se desviar dinamicamente para outros pontos do próprio documento ou para outro documento; conter conteúdo dinâmico, ou seja, gerado automaticamente a partir de bancos de dados; ou ter funções de navegação, que são barras de ferramentas, pesquisa a palavra-chave ou tabelas interativas de conteúdos;

VI *Preservar a autenticidade*: É necessário garantir que o objeto acessado seja exatamente aquele que se procura e que as possíveis transformações pelas quais tenha passado, para manter sua acessibilidade, preservaram sua forma original;

VII *Acompanhar o objeto digital ao longo do tempo*: Os objetos digitais tomam-se passíveis de serem alterados, copiados ou movimentados imediatamente após a sua criação. Em qualquer referência ao objeto digital, é necessário localizá-lo na edição ou versão correta;

VIII *Preservar a proveniência*: Identificar a origem do objeto e detalhar seu histórico ajuda a confirmar sua autenticidade e integridade;

IX *Preservar o contexto*: Os objetos digitais são definidos por sua dependência de hardware e software, seus modos de distribuição e seus relacionamentos com outros objetos digitais.

Da mesma forma, os fatores condicionantes foram cotejados com os nove requisitos da preservação digital apontados por Bullock (1999)<sup>13</sup>, obtendo-se a distribuição apresentada no GRAF. 4. Os resultados demonstraram que as maiores concentrações de fatores recaem, em ordem decrescente, sobre preservar o conteúdo com 75 fatores, preservar a presença física com 71 fatores, preservar a apresentação com 41 fatores e preservar a funcionalidade com 40 fatores.

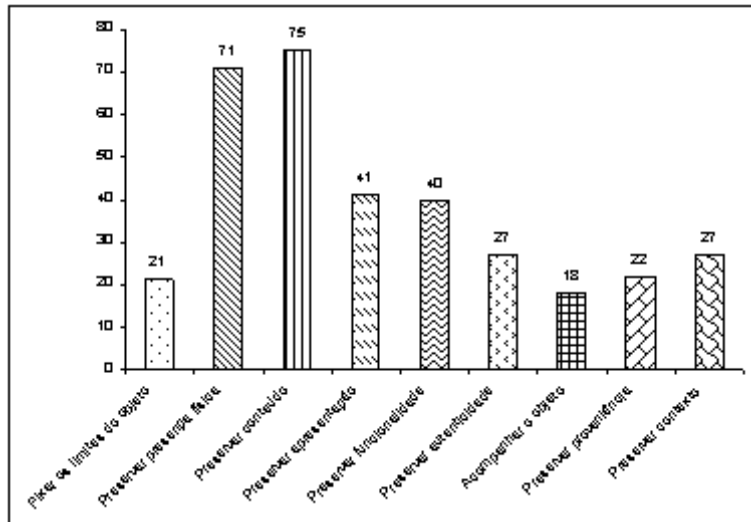


GRAFICO 4- Distribuição dos fetores condicionantes da preservação digital por requisito bullock

Finalmente, os fatores condicionantes foram confrontados com as seis funções arquivísticas indicadas no modelo de referência *Open archival information system - OAIS*<sup>14</sup> (CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS - CCSDS, 2002), constatando-se a distribuição apresentada no GRAF. 5, ou seja, 75 fatores na Administração do sistema, 50 fatores no Planejamento de preservação, 17 fatores no Arquivamento, 9 fatores no Acesso, 6 fatores na Admissão e 6 fatores no Gerenciamento de dados.

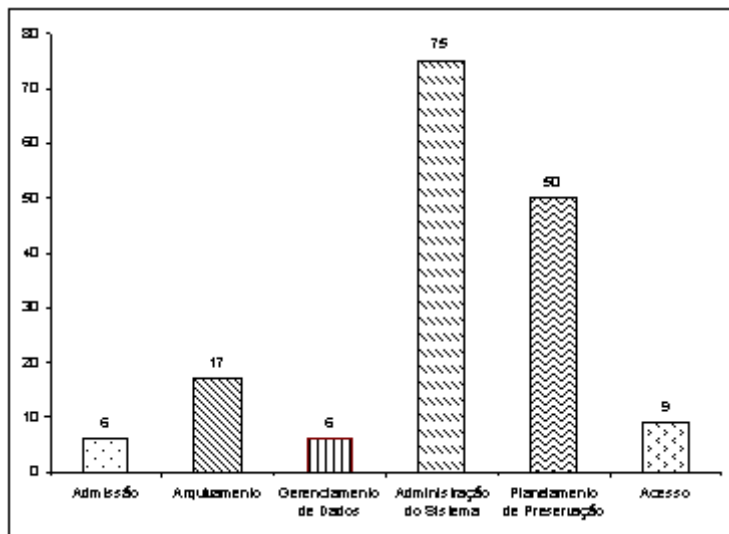


GRAFICO 5- Distribuição dos fetores condicionantes da preservação digital por funcao arquivística OAIS

Dessas comparações percebeu-se que, no diálogo entre classes-fatores condicionantes-literatura, novos conteúdos poderiam contribuir para a expansão ou detalhamento da FIG. 1. De fato, novas abordagens foram desenvolvidas, tendo culminado no modelo analítico do ambiente de preservação digital, elaborado de acordo com o diagrama de classe da *Unified Modeling Language - UML*<sup>15</sup>.

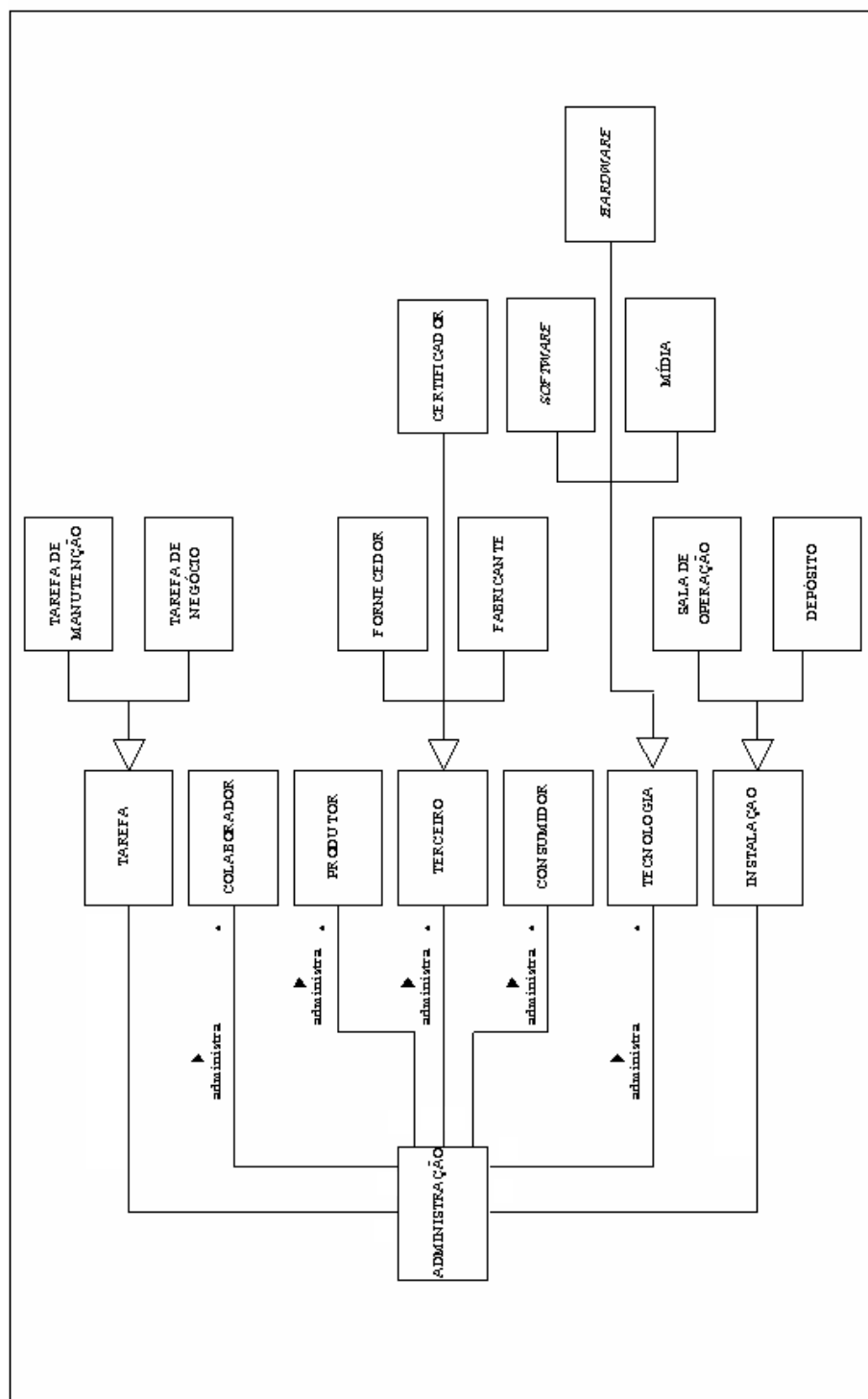


FIGURA 2a - Ambiente de preservação digital - modelo analítico

<sup>14</sup> O modelo de referência OAI, esquema conceitual que disciplina e direciona o sistema para a preservação e a manutenção do acesso à informação digital em longo prazo, identificou no primeiro nível de detalhamento, seis funções arquivísticas a serem exercidas:

I *Admissão (Ingest)*: Serviços e processos que aceitam pacotes de submissão de informação dos produtores, preparam pacotes de arquivamento de informação para arquivamento e garantem que esses pacotes de arquivamento de informação e sua informação descritiva de apoio sejam inseridos dentro do OAI;

II *Arquivamento (Archival Storage)*: Serviços e processos usados para o arquivamento e a recuperação dos pacotes de arquivamento de informação;

III *Gerenciamento de dados (Data Management)*: Serviços e processos para preencher, manter e acessar ampla variedade de informações. Alguns exemplos dessas informações são catálogos e inventários daquilo que pode ser recuperado na entidade arquivamento, algoritmos de processamento que podem ser executados sobre dados recuperados, estatísticas de acesso de consumidor, faturamento a consumidor, pedidos dirigidos a evento, controles de segurança e programações, políticas e procedimentos do OAI;

IV *Administração do sistema (Administration)*: Serviços e processos necessários para o controle das operações contínuas das outras funções do OAI;

V *Planejamento de preservação (Preservation Planning)*: Serviços e processos usados para monitorar o ambiente OAI e fornecer recomendações para assegurar que a informação armazenada permaneça acessível por longo prazo à comunidade usuária alvo, mesmo que o ambiente computacional original se torne obsoleto; e

VI *Acesso (Access)*: Serviços e processos que possibilitam aos consumidores visualizar os acervos de informação arquivada e serviços relacionados.

<sup>15</sup> A UML é uma linguagem padrão amplamente utilizada no âmbito da ciência da computação para especificar, visualizar, documentar e construir artefatos de um sistema. O diagrama de classe é uma estrutura lógica estática em superfície de duas dimensões, que mostra uma coleção de elementos declarativos de modelo. (FURLAN, 1998)



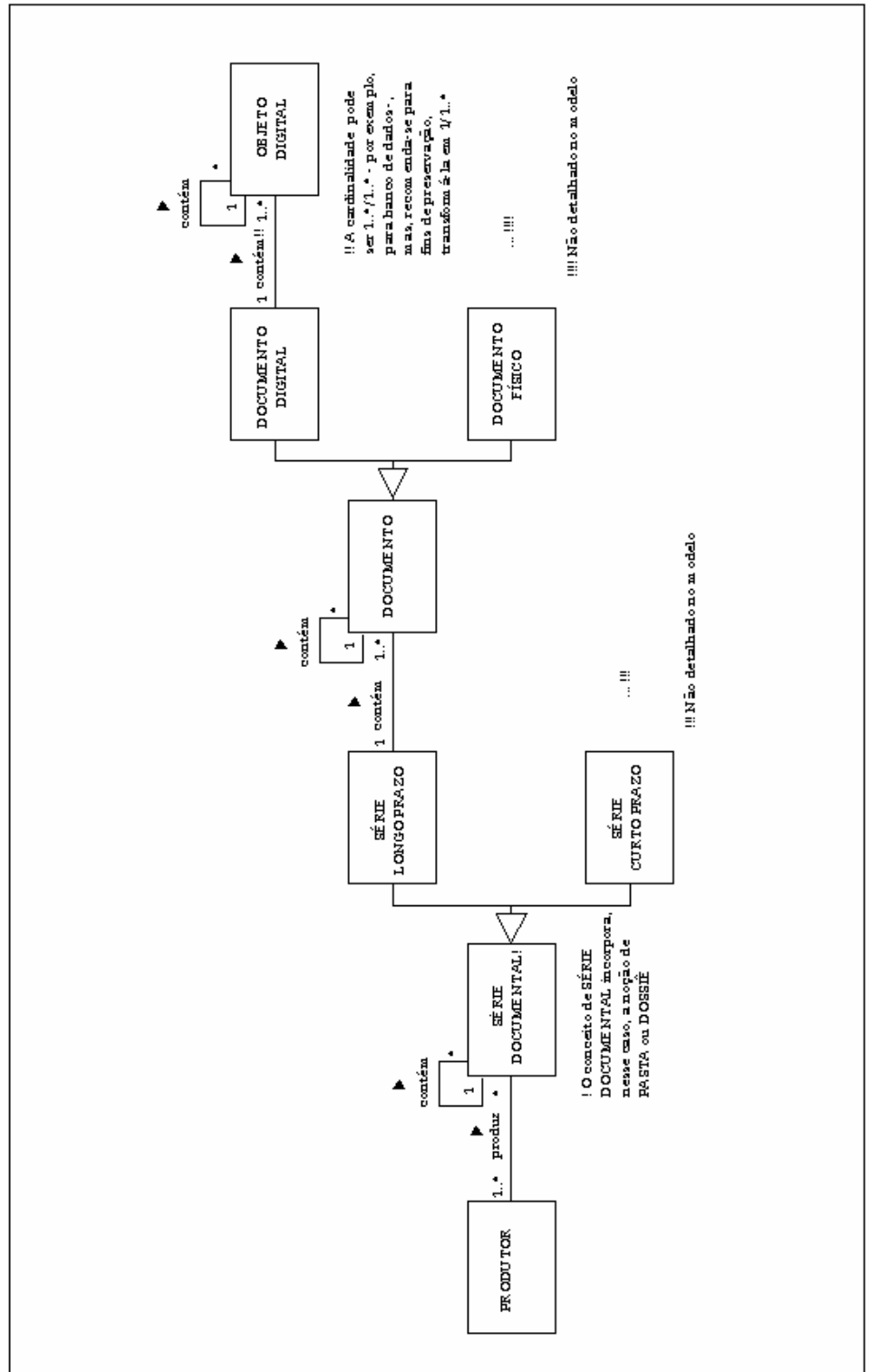


FIGURA 2b - Ambiente de preservação digital - modelo analítico (cont.)

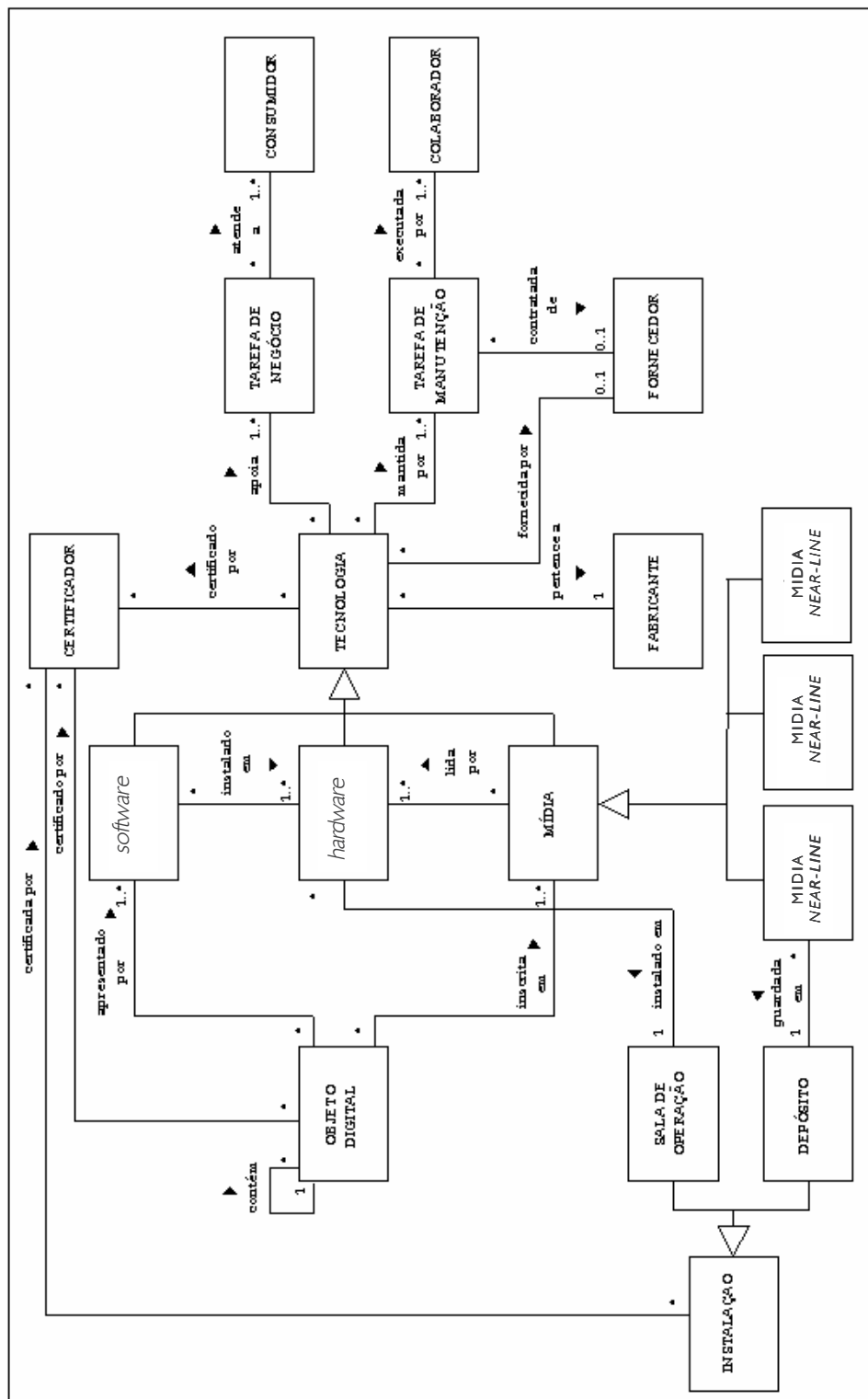


FIGURA 2c - Ambiente de preservação digital - modelo analítico (cont.)

O modelo analítico do ambiente da preservação digital desenvolvido na pesquisa encontra-se espelhado na FIG. 2a a 2d. (Os principais conceitos teóricos para leitura e entendimento do modelo podem ser obtidos no ANEXO B.) Procuraram retratar, além das relações entre as classes de objetos identificadas no ambiente de preservação digital, os desdobramentos – especializações – que determinam as diferentes abordagens e tratamentos. De fato, o modelo inicia com a administração gerenciando as demais variáveis organizacionais, cujos atos se traduzem na produção de séries documentais, que possuem documentos que podem apresentar valor permanente, dentre os quais aqueles em formato digital se constituem um ou vários objetos digitais depositados em determinada mídia de armazenamento e apresentados por *software* específico, ambos processados por *hardware* compatível. Esse conjunto tecnológico – mídia, *software* e *hardware* –, é mantido por tarefas de manutenção desempenhadas por colaboradores e apoia tarefas de negócio que atendem aos consumidores. O fluxo termina refletindo a influência do ambiente nas variáveis organizacionais, incluindo a própria administração, no sentido da mudança/evolução; e o ciclo se repete.

A inovação do modelo analítico oferece visão holística que permite ao profissional da informação e/ou ao pesquisador, análises mais precisas sem o esforço de busca de entendimento de cada caso. Qualquer situação de arquivamento digital se encontra espelhada no modelo, permitindo que cada aspecto apontado – classe, associação, especialização – seja objeto de implementação e estudos mais aprofundados, sem que o profissional ou o pesquisador esqueça algum componente ou perca a noção do todo. O principal objetivo do modelo analítico é proporcionar o entendimento mais amplo daquilo que seja necessário para preservar informação digital por longo prazo, tendo em vista as três categorias funcionais de metadados apontadas pelo *Working Group on Preservation Metadata* - PMWG (2001): descritivos, administrativos e estruturais<sup>16</sup>. Alguns pontos podem ser ressaltados. A passagem da série documental para a série de longo prazo (FIG. 2b) constitui um ponto. Outro, se refere à multiplicidade do documento de arquivo e à delimitação dos objetos de dados que compõem o referido documento (FIG. 2b). As tarefas de manutenção que, caso não reúnam informação suficiente sobre os componentes tecnológicos, poderão provocar a falência de todo o modelo (FIG. 2c). Pode-se mencionar ainda a influência do ambiente externo nas diversas classes (FIG. 2d). Associado ao modelo analítico, encontra-se o glossário de fatores condicionantes da preservação digital, que inclui amplo conjunto de termos e conceitos úteis desenvolvidos durante a elaboração da tese. O glossário considera também proposta de taxonomia – ordem, família, gênero e grau – destinada às tipologias de dados, mídia, *software* e *hardware*.

O modelo analítico reconhece a natureza altamente dispersa da informação digital na organização e a necessidade de implementações locais de políticas e procedimentos efetivos, através das diferentes variáveis organizacionais, que apoiem essa preservação. Considerou-se, no conjunto de classes envolvidas na preservação de informação digital de longo prazo, além da convencional tríade *hardware-software-dados*, o colaborador, o fornecedor, o fabricante, o certificador, a mídia, a sala de operação, o depósito, a administração e o ambiente externo. Essas classes trazem implícita, forte carga de informação que deverá ser considerada independentemente de sua posição no modelo e espelhada nos fatores condicionantes da preservação

<sup>16</sup> Segundo o PMWG (2001), Wendler observa que, no contexto dos objetos digitais, os metadados estão associados a uma de três categorias funcionais:

I *Descritivo*: facilita a descoberta e identificação de fonte;

II *Administrativo*: apoia o gerenciamento de fonte dentro de uma coleção;

III *Estrutural*: associa os componentes de objetos de informação complexos.

digital desenvolvidos na tese. Espera-se que qualquer organização com interesse no tema descubra que o uso dos termos e conceitos associados ao modelo analítico a auxiliará a atingir seus objetivos.

Resumindo, o modelo analítico e o glossário de fatores condicionantes da preservação digital oferecem esquema para:

- ◉ amplo entendimento e compreensão das variáveis organizacionais envolvidas na preservação da e no acesso à informação digital de longo prazo;
- ◉ padronização de terminologia e conceitos que permitam descrever e comparar abordagens de preservação existentes e futuras;
- ◉ possível expansão a outros esforços que incluam a preservação de informação em formato convencional (papel, microfilme etc.);
- ◉ possível identificação e produção de novos padrões.

O modelo analítico constante da FIG. 2a a FIG. 2d, o glossário de fatores condicionantes da preservação digital e as comparações com os pontos fundamentais da literatura da área representam abordagens inovadoras e originais e podem constituir excelentes ferramentas para a gestão da preservação digital, pois permitem avaliar os riscos associados às necessárias tomadas de decisão. Ao tomar a decisão, os fatores não considerados denunciariam automaticamente os riscos associados, isto é, as classes, problemas, requisitos e/ou funções não considerados ou debilmente considerados. Na ausência de informação existe apenas a incerteza. A informação transforma a incerteza em risco – a probabilidade estimada de que ocorrerá certo impacto negativo (GIDDENS, 1991). Com mais e melhores informações, o risco pode ser previsto de forma mais cuidada e ser melhor gerenciado. Representando o estado-da-arte atual, essas ferramentas devem ser constantemente revistas e atualizadas, adequando-se às novas soluções que venham a ser geradas pela evolução das tecnologias e processos.

## Algumas conclusões

A investigação desenvolveu-se num crescendo, em sucessivas *subinvestigações*, significando que, após as reflexões iniciais, explicitação do problema, escolha da metodologia, as conclusões foram se encaixando em direção ao *arredondamento* final. O valor da metodologia desenvolvida para estudar as mudanças tecnológicas e suas implicações para a preservação da informação digital reside em sua aplicação potencial em diferentes tipos de pesquisa na área da informação (arquivologia, computação, ciência da informação) dentre outras. De fato, a metodologia revelou-se extremamente eficaz para a concretização da pesquisa com o rigor desejável. Os instrumentos, no total de 14, foram gradativamente se refinando e dando origem a novos registros que apontavam para nova direção, culminando com o desenvolvimento do modelo analítico e do glossário de fatores condicionantes da preservação digital.

Cumpramos ressaltar que durante o desenrolar da pesquisa ocorreu a fundamental sensibilização dos profissionais das empresas pesquisadas (de grande porte), do Estado de Minas Gerais e do Município de Belo Horizonte, ainda que se tenha constatado não haver preocupação substancial relacionada ao tema. Merece destaque o envolvimento dos sujeitos pesquisados (pessoas extremamente competentes na sua área de atuação, que se entusiasmaram com a pesquisa, apresentando disponibilidade em todas as instâncias, o que contribuiu para o bom relacionamento pesquisador-respondente), que constituiu o ponto forte da pesquisa. Esse fato, certamente, propiciou o nível de excelência dos resultados alcançados.

A pesquisa passou da teoria à crítica aos princípios vigentes, bem como pela *praxis*, propondo-se ações de efetiva concretude. Esse fato ganha sentido em face da premência de pesquisas, discussões e práticas relacionadas aos temas do patrimônio e da memória.

### *Electronic records: factors serving as conditions for preservation*

*A doctoral dissertation, entitled Electronic records preservation: new challenges, old problems, was completed in the Doctor of Information Science Program of the School of Information Science, Federal University of Minas Gerais, Brazil, during 2000-2004, taking a transdisciplinary approach to the digital information archival environment in Brazilian public medium and large size organizations to identify specific factors serving as conditions for preservation. This article offers an overview of the investigation, sketching purposes and methodology, synthesizing the conceptual analysis and discussing the main discoveries..*

*Key-words: Archival science; Records management; Electronic records; Digital preservation*

## Referências

- BEAGRIE, N.; JONES, M. *Preservation management of digital materials: a handbook*. London: Digital Preservation Coalition, 2002. Disponível em < <http://www.dpconline.org/graphics/handbook/> > . Acesso em: 6 nov. 2003.
- BLAU, Peter M.; SCOTT, Richard. *Organizações formais*. São Paulo: Atlas, 1970.
- BOWDITCH, J. L.; BUONO, A. F. *Elementos de comportamento organizacional*. São Paulo: Pioneira, 1997.
- BRAND, S. Seeping the digital dark. *Library Journal*, v. 124, n. 2, p. 46-8, feb. 1999.
- BULLOCK, A. *Preservation of digital information: issues and current status*. Ottawa: National Library of Canada, 1999. Disponível em: < <http://www.nlc-bnc.ca/publications/1/p1-259-e.html> > . Acesso em: 25 nov. 2003.
- CHAMPION, D. J. *A sociologia das organizações*. São Paulo: Saraiva, 1985. 315 p.
- CHIAVENATO, I. *Teoria geral da administração*. 3.ed. São Paulo: McGraw Hill, 1987. v. 1.
- CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS. *Reference model for an Open Archival Information System (OAIS)*. Washington, 2002. 139p. Disponível em: < <http://www.ccsds.org/documents/650x0b1.pdf> > . Acesso em: 14 nov. 2002.
- CONWAY, P. *Preservation in the digital world*. Washington, DC: Commission on Preservation and Access - CPA, march 1996. Disponível em: < <http://www.clir.org/pubs/reports/conway2/> > . Acesso em: 23 nov. 2003.
- EUROPEAN COMMISSION, INTERCHANGE OF DATA BETWEEN ADMINISTRATIONS. *Model Requirements for the management of electronic records (MoReq)*. Luxembourg: European Commission, 2001. Disponível em: < <http://europa.eu.int/ISPO/ida/export/files/en/635.pdf> > . Acesso em: 23 out. 2003.
- FURLAN, J. D. *Modelagem de objetos através da UML: análise e desenho orientado a objetos*. São Paulo: Makron Books, 1998. 329 p.
- GIDDENS, Anthony. *As consequências da modernidade*. São Paulo: UNESP, 1991.
- HALL, R. H. *Organizações, estruturas e processos*. Rio de Janeiro: Prentice Hall do Brasil, 1984.
- HEDSTROM, M. Digital preservation: a time bomb for digital libraries. *Computer and the Humanities*, v.31, n.3, p.189-202, 1997/1998. Disponível em: < <http://www.uky.edu/~kiernan/DL/hedstrom.html> > . Acesso em: 25 nov. 2003.

- INTERNATIONAL COUNCIL ON ARCHIVES, COMMITTEE ON ELECTRONIC RECORDS. *Guide for managing electronic records from an archival perspective*. Paris, 1997. Disponível em: < <http://www.ica.org/biblio.php?pdocid=3> > . Acesso em: 20 fev. 2004. 58 p.
- LOSENET, Yola. Digital heritage for the future. *Cadernos DAD*, v. 2, p. 15-27, 2002.
- MIRANDA, G. Inácio M. P. *Organização e métodos*. 4. ed. São Paulo: Atlas, 1978.
- MADLER, D. A.; GERSTEIN, M. S.; SHAW, R. B. *Arquitetura organizacional*. Rio de Janeiro: Campus, 1993.
- NATIONAL LIBRARY OF AUSTRALIA. *Guidelines for the preservation of digital heritage*. Paris: UNESCO, 2003. 177 p. Disponível em: < <http://unesdoc.unesco.org/images/0013/001300/130071e.pdf> > . Acesso em: 25 nov. 2003.
- WORKING GROUP ON PRESERVATION METADATA. *Preservation metadata for digital objects: a review of the state of the art*. [S.l.]: Online Computer Library Center - OCLC, jan. 31, 2001. Disponível em: < [http://www.oclc.org/research/projects/pmwg/presmeta\\_wp.pdf](http://www.oclc.org/research/projects/pmwg/presmeta_wp.pdf) > . Acesso em: 6 dez. 2001.
- PUBLIC RECORD OFFICE. *Management, appraisal and preservation of electronic records: principles*. 2. ed. Kew: Public Record Office, 1999. Disponível em: < <http://www.pro.gov.uk/recordsmanagement/erecords/guidelines/principles.pdf> > . Acesso em: 19 fev. 2004.
- ROTHENBERG, J. *Ensuring the longevity of digital information*. Washington: Commission on Library and Information Resources. Disponível em: < <http://www.clir.org/pubs/archives/ensuring.pdf> > . Acesso em: 26 Nov. 2003. (Versão ampliada de Ensuring the longevity of digital documents. *Scientific American*, v. 272, n. 1, p. 42-47, jan. 1995).
- SANTOS, V. B. *Gestão de documentos eletrônicos: uma visão arquivística*. Brasília: Associação Brasileira de Arquivologia, 2002. 140 p.
- THIBODEAU, K. Building the archives of the future. *D-Lib Magazine*, v.7, n.2, p. 1-13, feb. 2001.
- WATERS, D.; GARRETT, J. *Preserving digital information: report of the Task Force on Archiving of Digital Information* commissioned by the CPA and the Research Libraries Group. Washington: Commission on Preservation and Access, 1996. Disponível em: < <http://www.rlg.org/ArchTF> > . Acesso em: 7 nov. 2003.

## ANEXO A - FATORES CONDICIONANTES DA PRESERVAÇÃO DIGITAL

CID CCT NOME

01 - Dados arquivados

01.01	01.32	identificação de dados
01.02	01.33	tipo de dados
01.03	01.30	localização de dados
01.04	01.23	função de dados
01.05	01.38	justificação de dados
01.06	01.12	sigilo de dados
01.07	01.11	temporalidade de dados
01.08	01.25	contexto de dados
01.09	01.39	acordos de dados
01.10	01.01	representação de dados
01.11	01.02	compactação de dados
01.12	01.22	prevenção de dados
01.13	01.31	certificação de dados
01.14	01.15	proteção lógica de dados
01.15	01.10	descrição de dados
01.16	01.06	apresentação de dados
01.17	01.37	semântica de dados
01.18	01.18	contingência de dados
01.19	01.26	responsabilidade por dados
01.20	01.27	área de dados
01.21	01.36	volume de dados
01.22	01.34	produção de dados
01.23	01.35	acesso a dados
01.24	01.24	procedência de dados
01.25	01.28	histórico de dados

02 - Mídia de armazenamento de dados arquivados

02.01	02.27	identificação de mídia de armazenamento de dados
02.02	02.28	tipo de mídia de armazenamento de dados
02.03	02.01	formatação de mídia de armazenamento de dados
02.04	02.24	localização de mídia de armazenamento de dados
02.05	02.25	certificação de mídia de armazenamento de dados
02.06	02.03	temporalidade de mídia de armazenamento de dados
02.07	02.26	requisitos de mídia de armazenamento de dados
02.08	02.30	restrições de mídia de armazenamento de dados
02.09	02.31	acordos de mídia de armazenamento de dados
02.10	02.06	descrição de mídia de armazenamento de dados
02.11	02.20	prevenção de mídia de armazenamento de dados
02.12	02.10	proteção física de mídia de armazenamento de dados
02.13	02.13	proteção lógica de mídia de armazenamento de dados
02.14	02.16	contingência de mídia de armazenamento de dados
02.15	02.21	descarte de mídia de armazenamento de dados

03 - *Software* de apresentação de dados arquivados

- 03.01 03.24 identificação de *software*
- 03.02 03.22 tipo de *software*
- 03.03 03.19 certificação de *software*
- 03.04 03.07 temporalidade de *software*
- 03.05 03.01 requisitos de *software*
- 03.06 03.23 restrições de *software*
- 03.07 03.21 acordos de *software*
- 03.08 03.04 interfaces de usuário
- 03.09 03.05 interfaces de *software*
- 03.10 03.06 descrição de *software*
- 03.11 03.11 proteção lógica de *software*
- 03.12 03.08 prevenção de *software*
- 03.13 03.14 contingência de *software*

04 - *Hardware* de processamento de dados arquivados

- 04.01 04.28 identificação de *hardware*
- 04.02 04.27 tipo de *hardware*
- 04.03 04.23 certificação de *hardware*
- 04.04 04.04 temporalidade de *hardware*
- 04.05 04.01 requisitos de *hardware*
- 04.06 04.25 restrições de *hardware*
- 04.07 04.05 interfaces de *hardware*
- 04.08 04.26 acordos de *hardware*
- 04.09 04.06 descrição de *hardware*
- 04.10 04.19 prevenção de *hardware*
- 04.11 04.10 proteção física de *hardware*
- 04.12 04.13 proteção lógica de *hardware*
- 04.13 04.16 contingência de *hardware*

## 05 - Tarefas de manutenção

- 05.01 05.35 identificação de tarefa de manutenção
- 05.02 05.03 descrição de tarefa de manutenção
- 05.03 05.29 acordos de manutenção
- 05.04 05.06 grau de eficácia de dados
- 05.05 05.30 grau de eficácia de prevenção de dados
- 05.06 05.34 grau de eficácia de certificação de dados
- 05.07 05.19 grau de eficácia de proteção lógica de dados
- 05.08 05.23 grau de eficácia de contingência de dados
- 05.09 05.09 grau de eficácia de mídia de armazenamento de dados
- 05.10 05.07 grau de ocupação de mídia de armazenamento de dados
- 05.11 05.31 grau de eficácia de prevenção de mídia de armazenamento de dados
- 05.12 05.17 grau de eficácia de proteção física de mídia armazenamento de dados
- 05.13 05.20 grau de eficácia de proteção lógica de mídia de armazenamento de dados
- 05.14 05.24 grau de eficácia de contingência de mídia de armazenamento de dados



- 05.15 05.27 grau de eficácia de descarte de mídia de armazenamento de dados
- 05.16 05.10 grau de eficácia de *software*
- 05.17 05.32 grau de eficácia de prevenção de *software*
- 05.18 05.21 grau de eficácia de proteção lógica de *software*
- 05.19 05.25 grau de eficácia de contingência de *software*
- 05.20 05.11 grau de eficácia de *hardware*
- 05.21 05.08 grau de ocupação de *hardware*
- 05.22 05.33 grau de eficácia de prevenção de *hardware*
- 05.23 05.18 grau de eficácia de proteção física de *hardware*
- 05.24 05.22 grau de eficácia de proteção lógica de *hardware*
- 05.25 05.26 grau de eficácia de contingência de *hardware*
- 05.26 05.12 grau de eficácia de colaborador
- 05.27 05.13 grau de eficácia de terceiro
- 05.28 05.40 grau de eficácia de instalação
- 05.29 05.36 grau de eficácia de condicionamento ambiental de instalação
- 05.30 05.37 grau de eficácia de prevenção de instalação
- 05.31 05.38 grau de eficácia de proteção física de instalação
- 05.32 05.04 custo de manutenção

06 - Tarefas de negócio

- 06.01 06.04 identificação de tarefa de negócio
- 06.02 06.03 requisitos de tarefa de negócio

07 - Colaboradores

- 07.01 07.07 identificação de colaborador
- 07.02 07.05 habilitações de colaborador
- 07.03 07.08 temporalidade de habilitações de colaborador

08 - Produtores de dados

- 08.01 08.01 identificação de produtor
- 08.02 08.02 habilitações de produtor
- 08.03 08.03 temporalidade de habilitações de produtor

09 - Consumidores de dados arquivados

- 09.01 09.07 identificação de consumidor
- 09.02 09.01 habilitações de consumidor
- 09.03 09.08 temporalidade de habilitações de consumidor
- 09.04 09.02 credenciais de segurança de consumidor
- 09.05 09.03 temporalidade de credenciais de segurança de consumidor
- 09.06 09.04 direitos de acesso de consumidor
- 09.07 09.05 temporalidade de direitos de acesso de consumidor

10 - Terceiros

- 10.01 10.05 identificação de terceiro
- 10.02 10.04 habilitações de terceiro
- 10.03 10.08 temporalidade de habilitações de terceiro

## 11 - Instalações

- 11.01 11.01 identificação de instalação
- 11.02 11.02 certificação de instalação
- 11.03 11.03 condicionamento ambiental de instalação
- 11.04 11.04 prevenção de instalação
- 11.05 11.05 proteção física de instalação

- 12.03 12.01 orçamento de TI
- 12.01 12.02 políticas e estratégias da organização
- 12.02 12.03 estrutura organizacional
- 12.03 12.01 orçamento de TI

## 13 - Ambiente

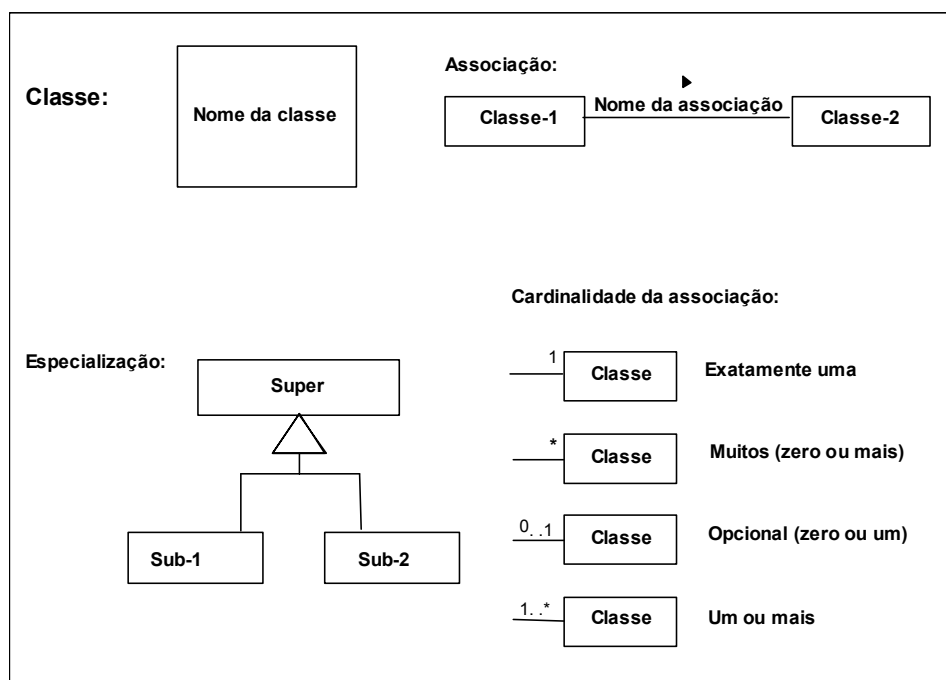
- 13.01 13.36 grau de relevância de dados
- 13.02 13.01 grau de confiabilidade de formatos de representação de dados
- 13.03 13.02 grau de confiabilidade de formatos de compactação de dados
- 13.04 13.38 grau de confiabilidade de prevenção de dados
- 13.05 13.42 grau de confiabilidade de certificação de dados
- 13.06 13.03 grau de confiabilidade de proteção lógica de dados
- 13.07 13.04 grau de confiabilidade de contingência de dados
- 13.08 13.05 grau de confiabilidade de mídia de armazenamento de dados
- 13.09 13.32 grau de confiabilidade de formatação de mídia de armazenamento de dados
- 13.10 13.39 grau de confiabilidade de prevenção de mídia de armazenamento de dados
- 13.11 13.07 grau de confiabilidade de proteção física de mídia de armazenamento de dados
- 13.12 13.08 grau de confiabilidade de proteção lógica de mídia de armazenamento de dados
- 13.13 13.09 grau de confiabilidade de contingência de mídia de armazenamento de dados
- 13.14 13.29 grau de confiabilidade de descarte de mídia de armazenamento de dados
- 13.15 13.10 grau de confiabilidade de *software*
- 13.16 13.13 grau de confiabilidade de interfaces de usuário
- 13.17 13.14 grau de confiabilidade de interfaces de *software*
- 13.18 13.40 grau de confiabilidade de prevenção de *software*
- 13.19 13.25 grau de confiabilidade de proteção lógica de *software*
- 13.20 13.26 grau de confiabilidade de contingência de *software*
- 13.21 13.11 grau de confiabilidade de sistema operacional
- 13.22 13.12 grau de confiabilidade de banco de dados
- 13.23 13.15 grau de confiabilidade de *hardware*
- 13.24 13.22 grau de confiabilidade de interfaces de *hardware*
- 13.25 13.41 grau de confiabilidade de prevenção de *hardware*
- 13.26 13.19 grau de confiabilidade de proteção física de *hardware*
- 13.27 13.20 grau de confiabilidade de proteção lógica de *hardware*
- 13.28 13.21 grau de confiabilidade de contingência de *hardware*
- 13.29 13.16 grau de confiabilidade de leitura de dados
- 13.30 13.17 grau de confiabilidade de apresentação de dados
- 13.31 13.23 grau de disponibilidade de especialistas de manutenção

- 13.32 13.34 grau de confiabilidade de terceiro
- 13.33 13.43 grau de confiabilidade de condicionamento ambiental de instalação
- 13.34 13.44 grau de confiabilidade de prevenção de instalação
- 13.35 13.45 grau de confiabilidade de proteção física de instalação
- 13.36 13.37 alertas de tecnologia

LEGENDA: CID - Código de identificação; CCT - Código de controle durante a pesquisa

## ANEXO B - CONCEITOS TEÓRICOS ACERCA DO DIAGRAMA DE CLASSE UML

A FIG. 3 apresenta de forma ilustrativa os elementos declarativos do diagrama de classe UML. Como se pode observar, uma classe é indicada por um retângulo contendo seu nome e representa um conjunto de objetos, isto é, coisas reais ou abstratas, reconhecidas como do mesmo tipo por compartilharem as mesmas características de atributos, operações, relações e semântica. As classes são relacionadas umas às outras através de associações, isto é, ligações entre classes. Uma linha sólida conectando duas classes demonstra a associação entre essas classes. A linha é identificada com o nome da associação, indicando a natureza da associação, e uma seta cheia apontando a direção que o relacionamento deve ser lido. A cardinalidade de cada classe é mostrada após a mesma, próximo à linha da associação. A cardinalidade refere-se à quantidade de instâncias da classe, isto é, à quantidade de objetos envolvidos no relacionamento .



Ainda com relação à FIG. 3, existem outros tipos de associações entre classes e uma delas é a especialização. A especialização é aquele tipo de associação, na qual a subclasse herda os atributos e métodos da superclasse. Na UML, um largo triângulo conectando a(s) subclasse(s) à superclasse demonstra esse tipo de associação. Uma instância da subclasse contém todos os atributos e métodos contidos em sua superclasse, de forma que a instância de subclasse possa ser usada em qualquer operação onde a instância de superclasse seja válida. A subclasse, entretanto, acrescenta novos atributos ou métodos, o que necessariamente impede que a subclasse seja seu substituto válido. Na FIG. 3, a especialização declara que os atributos e métodos da classe Super são herdados pelas classes Sub-1 e Sub-2.