

LUCIANO ANTONIO BERALDO

**A PRODUÇÃO DE INFORMAÇÃO EM CENTROS HISTÓRICOS
DE DOCUMENTAÇÃO: PROPOSTA DE GERENCIAMENTO
ELETRÔNICO DE DOCUMENTOS (GED) NO MUSEU PAULISTA**

São Paulo

2009

LUCIANO ANTONIO BERALDO

**A PRODUÇÃO DE INFORMAÇÃO EM CENTROS HISTÓRICOS
DE DOCUMENTAÇÃO: PROPOSTA DE GERENCIAMENTO
ELETRÔNICO DE DOCUMENTOS (GED) NO MUSEU PAULISTA**

Monografia apresentada à Escola Politécnica da
Universidade de São Paulo, como parte dos
requisitos para obtenção do título de MBA em
Tecnologia da Informação.

São Paulo
2009

LUCIANO ANTONIO BERALDO

**A PRODUÇÃO DE INFORMAÇÃO EM CENTROS HISTÓRICOS
DE DOCUMENTAÇÃO: PROPOSTA DE GERENCIAMENTO
ELETRÔNICO DE DOCUMENTOS (GED) NO MUSEU PAULISTA**

Monografia apresentada à Escola Politécnica da
Universidade de São Paulo, como parte dos
requisitos para obtenção do título de MBA em
Tecnologia da Informação.

Área de Concentração:
Tecnologia da Informação

Orientação: Prof. Dr. Alexandre Ricardo Nardi

Co-Orientação: Prof. Dra. Vera Lúcia N. Bittencourt

São Paulo

2009

AGRADECIMENTOS

A equipe do Museu Paulista da USP, que me apoiou em cursar o MBA e realizar esse trabalho.

A todos os professores e amigos do curso de MBA em TI, com quem tive o prazer de conviver durante dois anos.

A meus orientadores, pelas diretrizes e apoio durante todo meu trabalho.

RESUMO

O objetivo deste trabalho é propor uma metodologia para a implantação do Gerenciamento Eletrônico de Documentos (GED) no Museu Paulista da Universidade de São Paulo. Por possuir documentos de valor histórico inestimável, terão que ser adotados cuidados especiais na manipulação dos documentos. Durante esse trabalho são apresentados os conceitos e principais tecnologias para o entendimento do assunto. Uma análise do Serviço de Documentação Textual e Iconografia do Museu Paulista, local a ser informatizado, permitiu identificar elementos que justificam a utilização do GED e os benefícios a serem alcançados com a implantação. Durante este trabalho foi possível identificar as principais tecnologias a serem aplicadas. Através da aplicação dessa metodologia, buscou-se adiantar a resolução de questões necessárias a implantação do GED. Também foram levantadas diretrizes e normas visando auxiliar no processo decisório.

Palavras-chave: Gerenciamento Eletrônico de Documentos. Documentação. Tecnologia da Informação.

ABSTRACT

The goal of this work it is to propose a methodology to an implantation of Electronic Managing of documents (GED) in the Museu Paulista da Universidade de São Paulo. By having precious historical documents, it will have to be adopted special care in the manipulation of the documents. During this work concepts and main technologies are shown to the understanding of the subject. An analysis of the textual documentation job and iconography of the Museu Paulista; place witch will be computerized, allowed to identify the elements that justify the utilization of the GED and the benefits which will be achieved with the implantation. During this work it was possible to identify the main technologies to be applied. Trough the application of this methodology, sought after the resolution of the GED essential questions. Besides roles and routes were brought up to help in the decisive process.

Keywords: Electronic Managing of Documents. Documentation. Technology of Information.

LISTA DE ILUSTRAÇÕES

Figura 1 - Durabilidade das mídias (Mídias x Condições Ambientais x Durabilidade).....	34
Figura 2 - Museu Paulista da Universidade de São Paulo	36
Figura 3 - Museu Republicano “Convenção de Itu”	37
Figura 4 - Centro de Estudos do Museu, inaugurado em 2005.....	37
Figura 5 – Tela do sistema (módulo de iconografia)	41
Figura 6 - Problemas, Implicações e benefícios com a implantação do GED	45
Figura 7 - Fluxograma do processo.....	56
Figura 8 - Site José Bonifácio, Obra Completa	63
Figura 9 - Média de resoluções de câmeras digitais e dimensões ideais de impressão em papel fotográfico	66

LISTA DE ABREVIATURAS E SIGLAS

AIIM	The Association for Information and Image Management
CCS	Contínuos Composite Servo
CD-ROM	Compact Disc Read Only Memory
Cenadem	Centro Nacional da Gestão da Informação
CIA	Conselho Internacional de Arquivos
Cold	Computer Output to Laser Disk
Conarq	Conselho Nacional de Arquivos
CRM	Customer relationship management
CTDE	Câmera Técnica de Documentos Eletrônicos
DAM	Digital Asset Management
DI	Document Imaging
DM	Document Management
DPC	Pontos por centímetro
DPI	Dots per inch
DRM	Digital Rights Management
DVD	Digital Versatile Disk
ECM	Enterprise Content Management
EDMS	Engineering Document Management System
Embrapa	Empresa Brasileira de Pesquisa Agropecuária
ERM	Enterprise Report Management
ERP	Enterprise Resource Planning
EUA	Estados Unidos da América
Fogra	Associação para Pesquisa e Desenvolvimento da Tecnologia Gráfica
FTR	Full Text Retrieval
GED	Gerenciamento Eletrônico de Documentos
GIS	Geographic Information System
HD	Hard Disk
HD-DVD	High Density Digital Versatile Disc
ICR	Intelligent Character Recognition
IDM	Integrated Document System
IEC	International Electrotechnical Commission

IEEE	Institute of Electrical and Electronics Engineers
Iphan	Instituto do Patrimônio Histórico e Artístico Nacional
ISO	International Organization for Standardization
Jpeg	Joint Photographic Experts Group
LZW	Lempel-Ziv-Welch
Mbps	Megabits por segundo
MP	Museu Paulista
MRCI	Museu Republicano "Convenção de Itu"
OCR	Optical Character Recognition
OMR	Optical Mark Recognition
PDF	Portable Document Format
PHP	Hypertext Preprocessor
PMBOK	Project Management Body of Knowledge
PMI	Project Management Institute
RFP	Request for Proposal
RM	Records Management
SRGB	Standardised Red Green and Blue
SVDHICO	Serviço de Documentação Textual e Iconografia
TI	Tecnologia da Informação
Tiff	Tagged Image File Format
USP	Universidade de São Paulo
WCM	Web Content Management
WEB	World Wide Web
WFMC	Workflow Management Coalition

SUMÁRIO

INTRODUÇÃO	12
1 FUNDAMENTOS DO GERENCIAMENTO ELETRÔNICO DE DOCUMENTOS (GED)	16
1.1 A EVOLUÇÃO DO GED	16
1.2 CONCEITO DE GED	17
1.3 SIGLAS	19
1.4 APLICABILIDADES DO GED	20
1.5 IMPLANTAÇÃO DO GED NO BRASIL	23
1.6 TECNOLOGIAS QUE COMPÕEM O GED	24
1.6.1 Processo de Captura	24
1.6.1.1 Tecnologias relacionadas à Captura	26
1.6.2 Gerenciamento	29
1.6.2.1 Tecnologias Relacionadas ao Gerenciamento	29
1.6.3 Armazenamento	31
1.6.4 Distribuição	32
1.6.4.1 Tecnologias Relacionadas à Distribuição	32
1.6.5 Preservação	33
2 MUSEU PAULISTA DA USP	36
2.1 A INSTITUIÇÃO	36
2.2 MISSÕES ESSENCIAIS DE UM MUSEU UNIVERSITÁRIO	38
2.3 ACERVO DO MUSEU PAULISTA DA USP	39
2.4 ACERVO DO MUSEU REPUBLICANO CONVENÇÃO DE ITU	40
2.5 SISTEMA DOCUMENTAL INFORMATIZADO NO MUSEU PAULISTA	40

3 IMPLANTAÇÃO DE PROJETO DE GED NO MUSEU PAULISTA: METODOLOGIA, DIRETRIZES E EXPERIMENTAÇÃO	42
3.1 METODOLOGIA APLICADA	42
3.1.1 Descoberta	44
3.1.2 Levantamento / Análise	44
3.1.3 Projeto	55
3.1.4 Construção	59
3.2 DIRETRIZES E PADRÕES PARA OS SISTEMAS DE GERENCIAMENTO ELETRÔNICO DE DOCUMENTOS (GED)	59
3.2.1 Planejamento	59
3.2.2 Tecnologia de Compactação da imagem	60
3.2.3 Desenvolvimento da Requisição de Proposta (RFP - Request for Proposal)	60
3.2.4 Controle da Qualidade de Digitalização	60
3.2.5 Diretrizes quanto a Indexação	61
3.2.6 Campos que farão parte do sistema informatizado	61
3.2.7 Tecnologias de armazenamento	62
3.3 EXPERIMENTAÇÃO	63
4 CONCLUSÕES	71
REFERÊNCIAS	73
REFERÊNCIAS CONSULTADAS	76
ANEXO A – Glossário das Tecnologias	77

INTRODUÇÃO

O acesso ao conhecimento é entendido, na atualidade, como um dos principais componentes para a diminuição das desigualdades sociais. Por meio da democratização da informação cria-se conhecimento, gerando-se mão de obra qualificada e movimentos de desenvolvimento sócio-econômicos. Há de se considerar, portanto, a dimensão da tecnologia como condição para a disseminação do conhecimento, a configurar o que vem se definindo como “sociedade da informação”, onde a informação desempenha um papel fundamental na produção de riqueza e na contribuição para o bem-estar e qualidade de vida dos cidadãos.

Nesse sentido, os Museus, enquanto espaços a serviço da sociedade, devem contribuir para a socialização dos conhecimentos produzidos a partir de suas pesquisas oferecendo, para isso, diferentes canais de acesso aos seus diferentes públicos. Sensibilizada com essa necessidade, a Direção do Museu Paulista (MP) da Universidade de São Paulo (USP) solicitou o desenvolvimento de proposta de soluções tecnológicas, voltadas para tornar acessíveis os documentos existentes na instituição e, ao mesmo tempo, garantir sua preservação.

Assim sendo, este trabalho integra um conjunto de iniciativas voltadas para o estudo, desenvolvimento e implantação, de um Banco de Dados de amplo acesso destinado à pesquisa, tanto de especialistas como para o público em geral.

De início, a situação-problema com que o Museu se defronta pode ser resumida em: a) viabilizar o acesso aos acervos documentais da instituição, de forma a atender às demandas de democratização de informação e conhecimento; b) resgatar, preservar, organizar e tornar disponível acervo de grande relevância para a História do Brasil, ampliando fontes de pesquisa.

Por outro lado, constata-se que a busca pela otimização e agilidade das informações vem se constituindo em prioridade, especialmente para instituições culturais que têm como fim a preservação da memória e o estímulo à pesquisa, além da salvaguarda de documentos e coleções. Verifica-se que, a partir de 1990, um conjunto de iniciativas vêm se apresentando neste sentido, dialogando com políticas públicas definidas para o setor, emanadas do Ministério da Cultura e do Ministério da Ciência e Tecnologia. Essas políticas estão sustentadas em projetos e programas de

tecnologia de informação aplicados a acervos, tanto por equipamentos culturais públicos como o Museu Histórico Nacional, Biblioteca Nacional, Biblioteca Mário de Andrade, Museu do Índio, Museu Mineiro, como por meio de empresas gerenciadoras ou detentoras de acervos, como a Mameluco, responsável pelo processo de digitalização das obras do político brasileiro, José Bonifácio.

Nesse sentido, para que respostas às solicitações da sociedade possam ser atendidas em tempos aceitáveis para as atividades de pesquisa, a tecnologia da informação vem se constituindo em ferramenta adequada para auxiliar esta tarefa tão complexa. Assim, este trabalho propõe um estudo do Gerenciamento Eletrônico de Documentos (GED) enquanto parte integrante de projeto mais amplo de formatação de Banco de Dados, aberto a pesquisa em intranet e internet que assegure o acesso ao acervo, especialmente o textual, do Museu Paulista da USP.

Esse banco de dados também buscará interagir com os usuários, permitindo não só obter informações do acervo documental existente, mas ao mesmo tempo permitir que os próprios consulentes interajam com o sistema, permitindo a contribuição dos mesmos, através de inserção de elementos que possam contribuir com a geração do conhecimento. Este conhecimento será devolvido à sociedade através de exposições, livros, palestras, seminários, etc. Também será possível através de cadastro prévio para acesso ao sistema, obter o perfil do público consulente e contribuintes do sistema.

Trata-se, portanto de apresentar os benefícios da utilização do Gerenciamento Eletrônico de Documentos enquanto solução para a demanda proposta, assim como definir conceitos, funcionalidades, tecnologias, diretrizes, ações e padrões para a implantação de ferramentas relacionadas ao GED adequadas às exigências do Museu Paulista da USP e do público ao qual a instituição serve. Apoiando-se no estudo da bibliografia técnica sobre o GED, este trabalho levanta as seguintes hipóteses:

- a) A Tecnologia da Informação (TI) pode contribuir para a democratização de acesso a acervos documentais;
- b) O Gerenciamento Eletrônico de Documentos constitui condição para a formação e operacionalização de Banco de Dados;
- c) A atuação de analistas de sistemas, técnicos em informática e arquivistas representa diferencial importante para a eficiência de instituições voltadas

à salvaguarda de documentos e coleções, quanto à democratização de acesso e eficiência de mecanismos de pesquisa.

Cabe alertar que, o desenvolvimento da monografia voltou-se, também, para o estudo de casos e experiências visando encontrar sustentação teórico-metodológica para a proposta de gerenciamento eletrônico de documentos que estivessem relacionados com equipamentos culturais e acervos históricos.

Durante a elaboração das propostas do trabalho, foram necessárias consultas ao Serviço de Documentação Textual e Iconografia (SVDHICO) do MP a fim de coletar o maior número de informações referentes ao acervo documental, como por exemplo: a quantidade aproximada de documentos, sua forma de organização e armazenamento, como se dá o processo de consulta, o número de consultas aos arquivos e as principais dificuldades do serviço.

Na revisão da literatura procurou-se abordar diferentes temas como: tecnologias relacionadas ao GED, metodologias de implantação de projetos, normas e procedimentos para a digitalização de documentos, diretrizes e padrões para os sistemas de GED, etc. Além disso, estabeleceu-se contato com profissionais de empresas de digitalização de documentos e outras instituições relacionadas, procurando-se obter maiores informações sobre as tecnologias utilizadas e os processos para a implantação das mesmas, principalmente relacionados à documentação histórica.

Após o entendimento das tecnologias e processos utilizados para o gerenciamento eletrônico de documentos, foi possível buscar e propor a tecnologia adequada a cada etapa de desenvolvimento do projeto. Assim sendo, entendeu-se que o GED seria o instrumento mais adequado a ser avaliado em relação às necessidades do Museu Paulista.

Os capítulos desta monografia refletem este esforço de análise e experimentação propondo, primeiramente, um estudo bibliográfico sobre o GED. Em seguida, optou-se por apresentar o Museu Paulista de forma a facilitar o entendimento da situação problema. Nesta etapa, buscou-se conhecer a produção dos pesquisadores da instituição privilegiando-se o material descritivo ou de divulgação que, por seu aspecto sintético, forneceria ao leitor, as informações essenciais para o entendimento da instituição, que é alvo desse trabalho.

Através da aplicação de uma metodologia, procurou-se dar início ao desenvolvimento de um projeto piloto partindo da identificação dos problemas do Serviço de Documentação Textual e Iconografia, para que por meio destes e com o conhecimento bibliográfico adquirido fosse identificada a tecnologia do GED a ser aplicada. Também foram levantadas questões para nortear o caminho a ser seguido a fim de que a implantação do GED possa alcançar sucesso. Assim sendo, foi fundamental nesta etapa buscar diretrizes do setor Serviço de Documentação Textual e Iconográfica do MP para que sejam seguidos padrões adotados pela instituição.

1 FUNDAMENTOS DO GERENCIAMENTO ELETRÔNICO DE DOCUMENTOS (GED)

O Gerenciamento de Documentos, tanto os digitais como convencionais, na sociedade de informação vem se apresentando como um desafio que precisa ser considerado e superado. Considera-se imprescindível que a documentação de forma geral e, mais especificamente, aquela reunida em equipamentos culturais como museus, centros de documentação e arquivos, possa ser democratizada, enquanto informação, a partir do uso de novas tecnologias sugeridas e implementadas pelo desenvolvimento de novas mídias e pela evolução de tecnologias eletrônicas digitais. “O amparo às identidades culturais nos novos meios resultará em benefícios evidentes na forma de incrementos da atividade econômica em geral e desenvolvimento da cidadania” (TAKAHASHI, 2000, p 8).

Pesquisas desenvolvidas pela tecnologia da informação apontam para as possibilidades de acesso e organização destas documentações por meio de um sistema informatizado conhecido como Gerenciamento Eletrônico de Documentos (GED).

1.1 A EVOLUÇÃO DO GED

A tecnologia que antecedeu o GED foi o microfilme, inicialmente utilizado por instituições bancárias, para a microfilmagem de cheques. Estas perceberam que a microfilmagem seria uma solução capaz de resolver o problema de espaço, pois os aspectos legais exigiam o armazenamento dessa documentação a longo prazo. Essa tecnologia, ainda hoje, é utilizada por diversas empresas e instituições.

Stringher, afirma que “uma das principais razões para se microfilmarem, está exatamente na economia de espaço, pela eliminação dos arquivos convencionais”.

Na década de 60, a legislação começou a dar os primeiros passos para garantir que os documentos microfilmados fossem aceitos juridicamente e que, em algumas vezes, fosse possível o descarte do original em papel. A lei nº 5.433 de 8

de maio de 1968, elaborada pela Câmara dos Deputados para regulamentar a microfilmagem no Brasil, estabeleceu, em alguns de seus artigos (STRINGHER, 2003, p.29):

- 1) Que fica autorizada a microfilmagem de documentos particulares e oficiais;
- 2) Que os microfilmes, certidões, traslados e cópias passavam a ter, em juízo ou fora dele, o mesmo valor dos documentos originais;
- 3) Os documentos microfilmados poderão ser eliminados a critério da autoridade competente; e,
- 4) Que a eliminação do original far-se-ia mediante lavratura de termo por autoridade competente.

A tecnologia do GED chegou ao Brasil na década de 1980 por meio de cursos, palestras, congressos e seminários difundidos pelo Centro Nacional da Gestão da Informação (Cenadem). Segundo Antônio Paulo de Andrade e Silva, diretor do Cenadem “É bem verdade que, àquela época, digitalizar documentos e ficar livre de montanhas de papéis era novidade, e bem atraente” (SILVA, 2009).

Com o desenvolvimento da informática, no final do século XX se exigia uma maior rapidez na circulação da informação. A chegada da rede mundial fez com que os fornecedores tivessem de adequar seus softwares a essa nova plataforma. Levar as informações, se possível em tempo real, tornou-se uma necessidade, uma vantagem competitiva. O que era antes apenas digitalizar papéis passou a ser um grupo agregado de tecnologias. O conceito de GED precisou ser ampliado (ibidem).

Devido às tecnologias de informática constantemente em mudanças, criou-se então o sistema híbrido que agrega a gravação em meios digitais e o microfilme, criando um meio adicional para a preservação dos documentos.

1.2 CONCEITO DE GED

Para o entendimento do conceito de GED, inicialmente serão apresentadas algumas entidades nacionais e internacionais representativas do setor. O Centro Nacional da Gestão da Informação (Cenadem), o Conselho Nacional de Arquivos (Conarq) e *The Association for Information and Image Management (AIIM)*.

Compõem este quadro nacional, importantes referências que vêm de associações internacionais que buscam, em função das próprias características do processo de configuração da sociedade de informação, definir processos que se coadunem com movimentos de globalização.

O Cenadem tem como um de seus objetivos difundir as novas e emergentes tecnologias da informação e documentação que estão diretamente relacionadas com o complexo informacional gerado numa empresa e por essa recebida, processada e distribuída (CENADEM, 2008a).

Indo ao encontro destes questionamentos, o Conarq assumiu como finalidade definir a política nacional de arquivos públicos e privados, bem como exercer orientação normativa visando à gestão documental e a proteção especial aos documentos de arquivo. Visando atender a estes objetivos, a Câmara Técnica de Documentos Eletrônicos (CTDE), grupo de trabalho ligado ao Conarq, redigiu e elaborou em 2006, um modelo de requisitos para sistemas informatizados de gestão arquivística de documentos, o e-ARQ Brasil (CTDE, 2006).

A AIIM é uma organização internacional que tem como objetivo ajudar usuários a compreender os desafios associados à gestão de documentos, conteúdo, registros e processos empresariais. A AIIM foi fundada em 1943; é uma autoridade internacional em *Enterprise Content Management* (ECM) (AIIM, 2009).

A partir da apresentação das entidades, serão descritos seus conceitos em relação ao GED.

O Cenadem define o Gerenciamento Eletrônico de Documentos (GED) como um grupo de tecnologias, divididas em cinco funcionalidades básicas: captação, gerenciamento, armazenamento, distribuição e preservação. Dentro dessas funcionalidades, cada tecnologia tem uma função específica. Essas tecnologias, trabalhando reunidas ou isoladamente, promovem a organização de informações não-estruturadas (CENADEM, 2004).

Entende-se como informação não-estruturada aquela que não está armazenada em sistemas de informação, tal como documentos, mensagem de correio eletrônico, arquivos de texto, imagem ou som, planilhas, etc.

O e-ARQ Brasil define o GED como conjunto de tecnologias utilizadas para organização da informação não estruturada de um órgão ou entidade, que pode ser dividido nas seguintes funcionalidades: captura, gerenciamento, armazenamento e distribuição (CTDE, 2006).

uma ferramenta que através de um gráfico representa uma imagem. Tem por finalidade garantir que as imagens foram digitalizadas nos padrões corretos.

É aconselhável utilizar tarjas de controle nos documentos a serem digitalizados, a fim de avaliar se as cores estão corretas e para preservar a informação dimensional do original. Caso a imagem ou lote de imagens não esteja adequado às diretrizes estipuladas, deve-se solicitar uma nova digitalização.

Digitalização e indexação das Imagens:

O acervo textual do Museu Paulista deverá passar por um processo de digitalização e indexação. A indexação é o processo de inclusão no Banco de Dados de informações significativas retiradas do documento com a finalidade de localização. Ex. código, título, autor, data do documento, entre outras informações.

As imagens a serem convertidas para o meio digital podem ser produzidas na própria instituição ou fora dela. Tendo o Museu Paulista a obrigação de preservar documentos históricos únicos, de valor inestimável, não será cogitada a digitalização externa à instituição.

Essa digitalização poderá ser efetuada, por meio da utilização de recursos próprios (infra-estrutura tecnológica e mão de obra) ou de contratação de empresa terceirizada.

Em 2001, a Empresa Brasileira de Pesquisa Agropecuária (Embrapa) desenvolveu um Guia Geral de Digitalização de Documentos, visando fornecer orientações e recomendações sobre a digitalização dos documentos da empresa. Em Dez/2005, começou a ser produzida a nova versão com o objetivo de revisar e atualizar o documento original. Será utilizado esse guia, como referencial (Embrapa, 2006).

Recomenda-se que o Museu Paulista digitalize e indexe seus documentos, com a ajuda de contratação de empresa terceirizada, devido aos seguintes fatores:

- Não contar com recursos humanos suficientes;
- Não contar com a expertise necessária, que as empresas de digitalização possuem;
- Não possuir equipamentos em quantidade e adequados à digitalização em escala;

- Pagar pelo custo somente da digitalização da imagem, não pelo equipamento ou equipe;
- Poder digitalizar as imagens com alto nível de produção;
- Não arcar com custos relacionados à obsolescência da tecnologia, falhas, tempo de inatividade do sistema, necessidade de treinamento de mão-de-obra, etc.
- Necessidade de um tempo menor para a implementação do processo, da infra-estrutura técnica, e de geração de imagens;

Contudo, é necessário produzir um contrato detalhado, descrevendo todas as especificações necessárias, direitos e deveres de ambas as partes.

Desenvolvimento do Software:

O Museu Paulista deverá também adquirir soluções prontas no mercado, ou desenvolver seu próprio sistema informatizado, para o gerenciamento de seu acervo textual. Caso opte pelo desenvolvimento, o sistema poderá ser elaborado pela própria instituição ou através de empresa desenvolvedora a ser contratada.

Será tomada como base do processo decisório uma palestra apresentada por Antonio Geraldo da Rocha Vidal⁶, intitulada "Informatização nas Pequenas e Médias Empresas", realizada na Faculdade de Economia e Administração da USP (BOSTELMANN, 1991).

Recomendações acerca do desenvolvimento do software

O Museu Paulista conta com um único desenvolvedor, que também é o responsável pela Seção Técnica de Informática. Nas condições já indicadas, não há possibilidade de dedicação exclusiva do Analista para o desenvolvimento do *software*. Portanto, é aconselhável inicialmente a procura por empresas que tenham a solução pronta. Caso não seja encontrada a solução que se adapte à necessidade do Museu Paulista, opta-se pelo desenvolvimento específico.

⁶ Antonio Geraldo da Rocha Vidal é Doutor em Administração de Empresas pela Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo (FEA/USP), mestre em Administração de Empresas pela FEA/USP, Bacharel em Administração de Empresas pela FEA/USP e Engenheiro Civil pela Escola Politécnica da Universidade de São Paulo.

Os principais motivos que levam à recomendação de solução pronta são:

- Baixo custo do sistema e rapidez de implantação;
- Evita a manutenção de uma equipe de profissionais especializados na empresa;
- Garantia de que o sistema já foi suficientemente testado e funciona corretamente;
- Acompanhamento da evolução tecnológica da área por parte do fornecedor.

Caso não sejam encontradas soluções adequadas as necessidade do MP, recomenda-se a contratação de empresa terceirizada para o desenvolvimento específico. Os principais motivos que levam a essa recomendação são:

- Garantia de que todos os requisitos do sistema serão atendidos;
- Evita a manutenção de uma equipe de profissionais especializados na empresa;
- O sistema poderá acompanhar a dinâmica da empresa.

3.1.3 Projeto

Após a clara definição das necessidades, sugere-se inicialmente a execução de um projeto piloto para avaliação e melhoria da solução. Como ponto de partida, poderia ser utilizada a coleção ou coleções mais consultadas. Recomenda-se que o Museu Paulista crie um documento, com os requisitos essenciais e desejáveis.

Este documento deverá incluir os seguintes itens:

a) Descrição e Fluxograma do Processo

Deverá ser descrito todo o processo a ser implementado, acompanhado do fluxograma. A descrição do processo consiste em contextualizar os envolvidos a respeito do processo de implantação do GED. O Fluxograma permitirá a visualização do processo como um todo. O fluxograma e a descrição do processo serão apresentados abaixo:

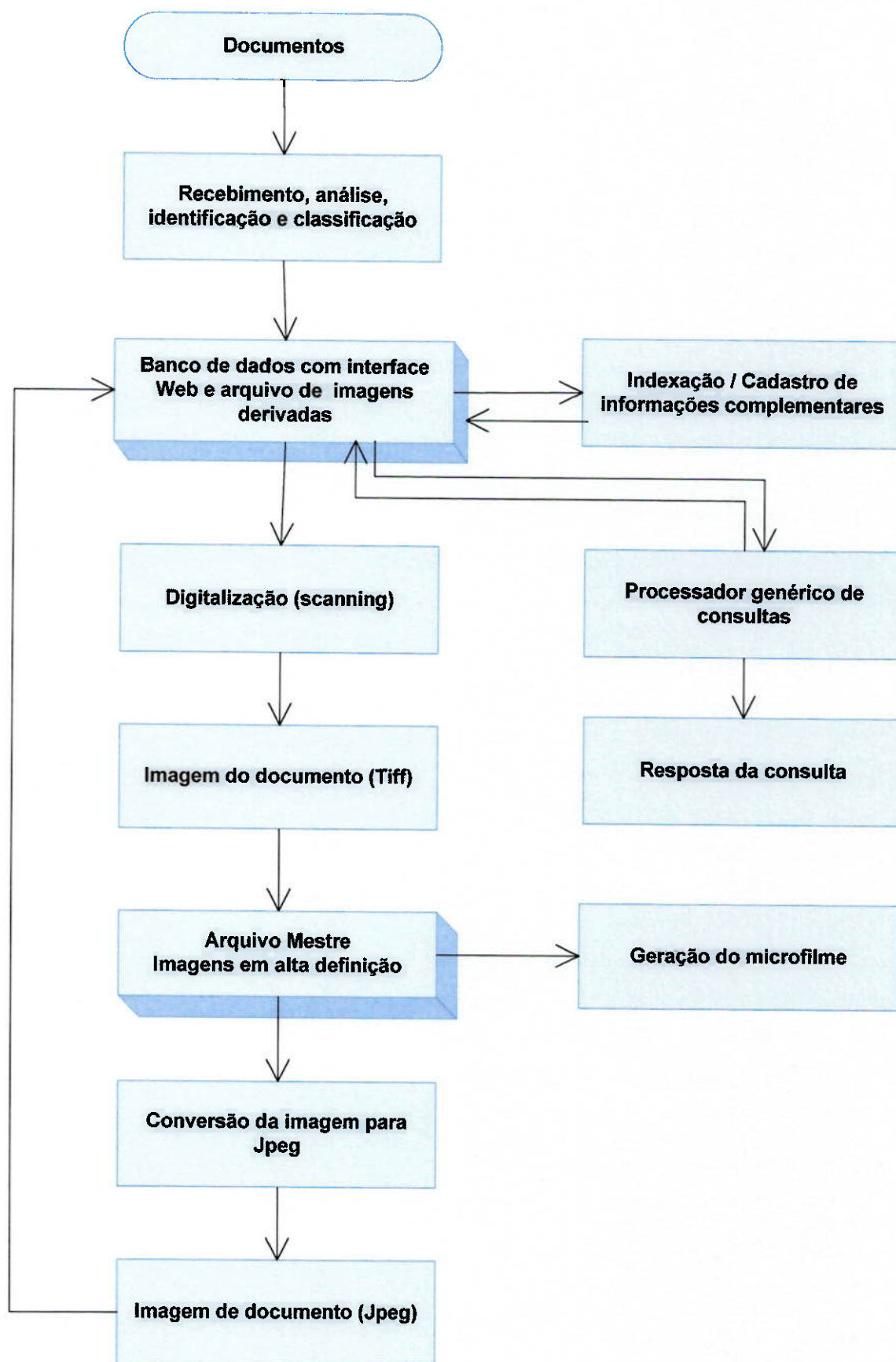


Figura 7 – Fluxograma do processo

Descrição do Processo:

O trabalho será iniciado pelo cadastramento do documento no Banco de Dados, através da atribuição de um código único de identificação.

Os documentos serão digitalizados em formato Tiff e serão armazenados em um servidor de imagens de alta resolução ao qual denominaremos de “Arquivo mestre”. A partir das imagens Tiff geradas, será criado o microfilme. Há equipamentos no mercado que geram o “filme” através de diversos formatos de imagem.

O arquivo Tiff, em seguida, será convertido, por meio de software, para o formato Jpeg e encaminhado para o servidor de Banco de Dados e imagens derivadas. A imagem será relacionada ao Banco de Dados através de seu código de identificação.

A qualquer momento, poderão ser cadastradas informações que permitam a localização do documento (indexação). Como o documento já está incorporado ao Banco de Dados, o mesmo poderá ser utilizado como referência.

Também poderão ser inseridas informações complementares, tais como: fonte, data da aquisição, valor em dólar, valor em reais, avaliador, data da última avaliação, circulação, histórico, descritores, entre outras.

b) Especificação dos requisitos de digitalização

O setor de documentação textual e iconografia e a seção técnica de informática devem descrever os procedimentos a serem seguidos referentes à digitalização do acervo. Esse documento deverá conter entre outras coisas:

- Descrição e fluxo de trabalho quanto à digitalização;
- Procedimentos e cuidados no processo de digitalização;
- Processo do manuseio dos originais;
- Aspectos referentes aos formatos de gravação e resolução da imagem digital;
- Procedimentos para avaliação da qualidade da imagem gerada;
- Produção diária a ser cumprida.

c) Especificação dos requisitos de software

Os requisitos para o sistema de Gerenciamento do Acervo Textual do MP deverão ser descritos, entre eles: os de captura, localização e apresentação dos documentos, segurança, armazenamento, preservação, funções administrativas, usabilidade, disponibilidade, desempenho, escalabilidade, entre outros.

O e-ARQ Brasil, documento elaborado pela Câmara Técnica de Documentos Eletrônicos, estabelece os requisitos mínimos para um Sistema Informatizado de Gestão Arquivística de Documentos, independentemente da plataforma a ser utilizada. O modelo deverá ser utilizado como base para a definição dos requisitos de software a serem aplicados no Museu Paulista (CTDE, 2006).

d) Especificações dos requisitos de *Hardware* e infra-estrutura

Deverão ser especificados detalhadamente os requisitos de hardware, entre eles: especificações dos servidores, estações de trabalho, impressoras e infra-estrutura necessária à implantação do projeto.

e) Especificações dos Recursos Humanos necessários

Um sistema eficiente de documentação exige uma equipe conhecedora dos problemas de informação principalmente quanto à armazenagem e recuperação dessas informações. Deve-se ter claro que a documentação em si não é o objetivo principal, mas a obtenção da informação na hora em que se precisa, de forma ágil e precisa.

Portanto, deverão ser especificados os recursos humanos e suas habilidades para a implantação do projeto, quanto ao processo de digitalização, indexação, cadastramento das informações complementares e gerência do projeto.

f) Estimativa de Custo e Cronograma

A estimativa de custo e cronograma é a última etapa a ser executada. Pressupõe-se que todas as informações necessárias para a execução do projeto já tenham sido levantadas.

3.1.4 Construção

Nesta fase, é feita a implementação do projeto, objetivando a instalação de hardware e compra de software ou desenvolvimento, monitoramento do processo e treinamento aos usuários.

3.2 DIRETRIZES E PADRÕES PARA OS SISTEMAS DE GERENCIAMENTO ELETRÔNICO DE DOCUMENTOS (GED)

Para o desenvolvimento do Projeto Piloto, deverão ser seguidas as diretrizes e padrões já consagrados pelos diferentes setores envolvidos no processo, que irão ajudar o Museu Paulista no êxito de sua implementação.

Essas práticas deverão ser especificadas em documentos de aquisições e contratos para que os fornecedores sejam obrigados a seguir um padrão adequado. As diretrizes irão fornecer informações que ajudarão a preparar, selecionar e implementar, com êxito, a tecnologia selecionada (AIIM, 2008).

3.2.1 Planejamento

O planejamento envolve o levantamento e a análise da realidade institucional, o estabelecimento das diretrizes e os procedimentos a serem cumpridos pelo órgão ou entidade, o desenho do sistema de gestão arquivística de documentos e a elaboração de instrumentos e manuais.

Diretrizes quanto ao planejamento para implementação do DI e a preparação dos documentos a serem digitalizados, podem ser encontradas em:

- ISO/DTR 12652 – Relatório técnico sobre a preparação dos documentos existentes para os sistemas de captação de imagem;
- ANSI/AIIM TR15-1997 – Considerações sobre Planejamento, Abordando a Preparação de Documentos para a Captação da Imagem.

3.2.2 Tecnologia de compactação de imagem

Diretrizes que habilitam a instituição a selecionar a tecnologia de compactação mais adequada podem ser obtidas através das normas:

- ISO/TS 12022:2001 – Orientação sobre métodos de compactação da imagem dos documentos.
- ISO 12033 - Guia de métodos de compressão para imagem de documentos.
- ANSI / AIIM TR33-1998 – Seleção de um Método de Compactação de Imagem Adequado para atender os Requisitos do Usuário.

3.2.3 Desenvolvimento da Requisição de Proposta (RFP - *Request for Proposal*)

Antes de selecionar um integrador, o Museu Paulista deverá enviar aos fornecedores um documento com os requisitos necessários do produto a ser adquirido. As normas quanto à elaboração desse documento podem ser encontradas em: ANSI/AIIM TR27-1996 – Diretrizes para RFP – Requisição de Proposta para Imagem Eletrônica.

3.2.4 Controle da Qualidade de Digitalização

As normas referentes à qualidade de digitalização dos documentos digitais podem ser obtidas em: ANSI / AIIM TR34-1996 – Procedimentos de amostragem para Inspeção por meio de Atributos de Imagens em Sistemas de Gerenciamento da Imagem de Documentos e Sistemas Micrográficos.

3.2.5 Diretrizes quanto a Indexação

O setor de documentação textual e iconografia deve estabelecer diretrizes e procedimentos aplicáveis aos campos de indexação para implantação do *Document Imaging* (DI). Essas diretrizes devem ser levadas em consideração quando do planejamento dos requisitos de indexação.

Cabe aos profissionais de arquivo do MP elaborar um thesaurus ou vocabulário controlado ou ainda lista de termos (descritores), normalizado pela instituição, que servirá para indexação e recuperação dos documentos.

A diretriz ANSI/AIIM TR40-1995 sugere campos de índice para Documentos em Ambientes de Imagem Eletrônica.

3.2.6 Campos que farão parte do sistema informatizado

Deverá ser utilizada a descrição arquivística para seleção de campos que farão parte do sistema informatizado para a Gestão de Documentos. Segundo o Conselho Internacional de Arquivos (CIA), o objetivo da descrição arquivística é identificar e explicar o contexto e o conteúdo de documentos a fim de promover o acesso aos mesmos. Foi elaborado pela CIA, o ISAD(G)⁷, que fornece as normas gerais internacionais de descrição arquivística.

As regras estão organizadas em sete áreas de informação descritiva:

- Área de Identificação (destinada à informação essencial para identificar a unidade de descrição);
- Área de contextualização (destinada à informação sobre a origem e custódia da unidade de descrição);
- Área de conteúdo e estrutura (destinada à informação sobre o assunto e organização da unidade de descrição);

⁷O ISAD(G) é uma norma geral internacional de descrição arquivística. Ver em: <http://www.conarq.arquivonacional.gov.br/Media/publicacoes/isad_g_2001.pdf>. Acesso em: 10 jan. 2009.

- Área de condições de acesso e de uso (destinada à informação sobre a acessibilidade da unidade de descrição);
- Área de fontes relacionadas (destinada à informação sobre fontes com uma relação importante com a unidade de descrição);
- Área de notas (destinada à informação especializada ou a qualquer outra informação que não possa ser incluída em nenhuma das outras áreas);
- Área de controle da descrição (destinada à informação sobre como, quando e por quem a descrição arquivística foi elaborada).

3.2.7 Tecnologias de armazenamento

Durante o estágio de planejamento do projeto, a organização deve analisar as diretrizes para o planejamento, implementação e operação de armazenamento de longo prazo e permanente de registros públicos em mídia eletrônica.

Algumas normas referentes ao armazenamento podem ser obtidas em:

Relatórios Técnicos de Armazenamento e Arquivamento

ANSI/AIIM TR41-2006 – Tecnologia, Gerenciamento e Padrões para armazenamento em disco óptico.

Padrões para Armazenamento e Arquivamento Óptico com *Blue Laser*

ISO/IEC 17345:2004 – Tecnologia da Informação – Intercâmbio de dados em cartuchos de disco UDO (ultra densidade óptica) regravável e gravável de 130 mm – Capacidade: 30 Gbytes por cartucho (primeira geração).

Padrões para Armazenamento e Arquivamento com *Red Laser*

- ANSI INCITS 212-1992 (R2002) (ISSO/IEC 10089:1991) – Cartucho de Disco óptico Regravável de 130mm para Intercâmbio de Informações.
- ISO/IEC 15286:1999 – Cartucho de disco óptico de 130 mm utilizando o efeito magneto-óptico para funcionalidade *write once, ready many*.
- ANSI INCITS 234-1993 (R2003) – Métodos de Teste para características de mídia Cartuchos de disco óptico regravável de 130mm para armazenamento de dados com *Contínuos Composite Servo (CCS)*.

3.3 EXPERIMENTAÇÃO

Segundo BASTOS (2009), o projeto José Bonifácio, Obra Completa, cujo nome identifica o seu objetivo principal, foi desenvolvido como um modelo de operação multidisciplinar, onde especialistas em suas respectivas áreas de atuação atuaram em conjunto no planejamento, na capacitação de pessoal, na implantação e na execução das tarefas e dos serviços necessários ao seu objetivo fundamental, ou seja, tornar acessíveis as obras e documentos relacionados ao personagem.

Esse projeto foi uma iniciativa com financiamento privado junto às instituições públicas, detentoras de acervos e/ou documentos pertencidos ao personagem ou a ele diretamente ligados. O projeto visou localizar esses documentos e obras, tornando-os disponíveis para acesso tanto ao grande público quanto aos profissionais especializados, através do sítio de internet www.obrabonifacio.com.br.



Figura 8 – Site José Bonifácio, Obra Completa

Instituições como a Biblioteca Nacional, Biblioteca Mário de Andrade, Museu Paulista da USP, Museu de História Diplomática do Itamaraty, Arquivo Nacional, Instituto Histórico e Geográfico Brasileiro e o Museu Imperial de Petrópolis participaram do projeto.

O planejamento cuidadoso de cada uma das etapas do projeto foi o ponto determinante para o sucesso desse empreendimento, na avaliação de seus executores. Foi durante o planejamento que todas as questões levantadas pelos especialistas das diversas áreas envolvidas foram exploradas, buscando e eliminando as falhas que pudessem comprometer tanto o resultado dos esforços, quanto a segurança dos documentos originais.

A seguir, serão descritos os critérios e os procedimentos adotados na realização do projeto José Bonifácio, Obra Completa.

Iniciando com a estimativa da existência de aproximadamente 100.000 páginas de documentos e de livros raros a serem digitalizados no âmbito do projeto, definiram-se os parâmetros básicos pelos quais as operações envolvidas deviam funcionar. Um índice de acerto de 99,9%, apesar de parecer elevado, implicaria em colocar em risco ou causar dano à integridade de 1.000 páginas de documentos, podendo prejudicar seriamente a memória pública. Sob o ponto de vista da equipe de planejamento e da missão primordial do projeto, tal fato foi tratado como inaceitável. A margem de erro deveria ser reduzida para o menor valor possível.

Assim, definiu-se que a fase de planejamento, prevista inicialmente para ter curta duração, deveria ser estendida, afim de que se pudesse explorar adequadamente as etapas e os processos envolvidos nesta empreitada.

O planejamento foi baseado em dois pressupostos fundamentais:

- Não causar, em hipótese alguma, dano aos originais, expondo-os ao processo apenas o mínimo necessário para a sua correta identificação, registro e digitalização.
- Registrar a maior quantidade de informações numa planilha de coleta, atrelada ao número de inventário adotado. Entre essas informações constam: identificação do documento, título, estado de conservação, primeiro parágrafo do documento para que não haja dúvida quanto sua identificação.

O projeto foi então desenhado, sendo cada uma de suas fases estudadas detalhadamente para atender aos requisitos propostos. Definiu-se então que os

critérios operacionais de cada uma delas deveriam ser baseados em padrões amplamente reconhecidos e aceitos internacionalmente, nas diferentes áreas de atuação que compunham cada etapa.

Foi desenvolvido o fluxograma operacional contendo todas as etapas de processos envolvidas, da definição das estratégias iniciais, passando pela negociação de termos de parcerias com as instituições de guarda, pelo processo de coleta de dados, pela digitalização e pelo pós-processamento de toda a informação adquirida até a sua publicação final no sítio de internet do projeto.

Procedimentos rigorosos foram definidos para nortear a escolha e a atuação dos envolvidos, fossem eles pessoas ou empresas. Elaboraram-se desde regras de conduta pessoal dentro das instituições envolvidas, de asseio e de organização, normas específicas de como lidar com os originais (baseados também nos padrões das mais renomadas instituições mundiais), bem como a adoção de elevados critérios técnicos para a correta escolha das tecnologias e dos equipamentos mais adequados.

Manuais de procedimentos específicos a cada etapa foram elaborados e *workshops* de capacitação foram ministrados às equipes, pelos especialistas das áreas envolvidas.

Na definição dos procedimentos de digitalização, foram adotados os padrões de reprodução de imagens e cores definidos pela *Commission Internationale de l'Eclairage* (Comissão Internacional de Iluminação), com sede em Viena, Áustria, entidade que visa discutir e padronizar os assuntos relativos à ciência e à arte da iluminação, das cores, reprodução e transmissão de imagens, além das recomendações dos principais institutos de normalização, como a Associação para Pesquisa e Desenvolvimento da Tecnologia Gráfica (Fogra) e *International Organization for Standardization* (ISO).

Definiu-se, como condicionante, que as digitalizações deveriam ser feitas em resolução adequada à formação de matrizes digitais com vida útil de longa duração, que poderiam dar origem a diversos tipos de arquivos digitais ou ainda, como salvaguarda digital ao original. Nesse ponto, os conhecimentos dos especialistas em reprodução gráfica foram essenciais.

Como a resolução de um documento digital de imagem é definida por uma matriz de pixels e uma matriz de resolução de saída (impressão de livros, jornais, cartazes, rascunhos, visualização em monitores, televisão, entre inúmeros outros), o

fator limitador encontra-se na adaptação entre ambas as matrizes. Assim, uma resolução de entrada baixa limita as dimensões de saída em um dispositivo de maior resolução, ao passo que uma resolução de entrada maior permite uma ampliação das dimensões na saída.

A seguinte regra tomou forma: a resolução de entrada ou de digitalização é equivalente à resolução do dispositivo de saída nas mesmas dimensões.

Utilizando scanner, a definição da resolução é uma tarefa simples; contudo ao se pensar em máquinas digitais essa tarefa não é tão simples assim. A resolução é medida em megapixels (número de *pixels* da largura x número de *pixels* da altura). Por exemplo: Para um original de tamanho 21 x 29,7 cm (A4), a ser fotografado com uma resolução de 300 dpi (ponto por polegada) e sendo uma polegada igual a 2,54 cm, será efetuado o seguinte cálculo:

$$300 \text{ dpi} \div 2,54 \text{ cm} = 118 \text{ dpc (pontos por centímetro)}$$

$$21 \text{ cm} \times 118 \text{ dpc} = 2478 \text{ pixels};$$

$$29,7 \text{ cm} \times 118 \text{ dpc} = 3504 \text{ pixels};$$

$$2478 \times 3504 = 8.682.912 \text{ pixels ou } 8,7 \text{ megapixels}$$

Tomando como base o cálculo acima, adotou-se a seguinte tabela aproximada de dimensões e resoluções de impressão de câmeras digitais.

Resoluções de câmeras digitais	Tamanho da imagem	Dimensões ideais de impressão (300 dpi).
640 X 480	0,3 <i>megapixel</i>	5,4 x 4,1 cm
1152 X 864	1 <i>megapixel</i>	9,8 x 7,2 cm
1280 X 960	1,2 <i>megapixel</i>	10,8 x 8,1 cm
1600 X 1200	2 <i>megapixels</i>	13,6 x 10,2 cm
2048 X 1536	3 <i>megapixels</i>	17,4 x 13 cm
2272 X 1704	4 <i>megapixels</i>	19,3 x 14,4 cm
2592 X 1944	5 <i>megapixels</i>	22 x 16,5 cm
3008 X 2000	6 <i>megapixels</i>	25,5 x 17 cm
3264 X 2448	8 <i>megapixels</i>	27,7 x 20,8 cm

Figura 9: Média de resoluções de câmeras digitais e dimensões ideais de impressão em papel fotográfico.

Sabendo que a grande maioria dos documentos tinha dimensões menores ou iguais às do formato A4 (210x297mm), chegou-se à conclusão de que a digitalização desses documentos com resolução acima de 9 megapixels era praticamente desnecessária.

Obteve-se assim uma operação de alto nível técnico, elevada eficiência, segura e com custos significativamente baixos por unidade digitalizada.

Em virtualmente 100% dos casos, os documentos foram digitalizados corretamente da primeira vez, em resolução adequada, evitando assim, constantes indas e vindas dos originais de seus repositórios.

Outro problema que poderia ser encontrado é o da representação tonal adequada do documento digital em relação ao original.

Como os dispositivos de captação são produzidos por processos industriais de grande escala (mesmo os profissionais), há uma relativa margem de erro considerada e tolerada em sua fabricação. Por isso, dois televisores de um mesmo fabricante não apresentam imagens idênticas em uma loja de eletrodomésticos.

A adoção de um espaço de cor padronizado torna-se extremamente importante. Adotou-se o *Standardised Red Green and Blue* (SRGB), instrução normativa ISO IEC 61966-2.1, espaço de cor definido pela ISO como referência universal para todos os sistemas de reprodução de cores. Todo e qualquer dispositivo, tanto de entrada, quanto de saída, é capaz de mapear seu espaço de cor nativo para o espaço SRGB.

Ainda assim, as variáveis ambientais podem alterar significativamente o resultado captado pelo dispositivo de digitalização.

Poder-se-ia argumentar que um técnico treinado poderia fazer a correção dos arquivos em momento posterior, mas, uma questão fundamental vem à tona, o correto procedimento em reprodução implica, necessariamente, em eliminar fatores subjetivos ou não mensuráveis. Para tanto, a adoção de escalas padronizadas é a única solução viável. Com a sua utilização, obtém-se o registro do momento da digitalização, que pode ser medido pelos mais simples programas de visualização de imagens, sem a necessidade da interpretação visual de um técnico (baseada em seu julgamento e não em fatores objetivos), sendo possível avaliar a exatidão do processo, mesmo se os monitores usados para esse fim não estiverem corretamente calibrados.

Outro ponto fundamental foi à opção pelo formato de gravação mais adequado ao cumprimento da missão do projeto, sem onerar os sistemas de armazenamento e de transporte de dados. Optou-se pelo formato Jpeg por este ser aberto e de domínio público, definido pelo *Joint Photographers Expert Group*, que vem sendo constantemente atualizado ao longo dos últimos anos. Apesar de adotar um sistema de compressão de informações destrutivo, observações em diversos sistemas de reprodução, utilizando-se compactação através de baixa compressão, demonstram claramente que essa destruição não é significativa ao olho humano, até uma ampliação média de 400x do tamanho original.

Outro fator considerado é que, por causa das leis naturais, há mais perda de informação em qualquer processo de transposição de uma imagem digital para o meio físico do que aquela gerada pelo algoritmo Jpeg, em suas menores taxas de compressão.

Isso resultou em poder digitalizar, com extrema precisão, quase 800 páginas de documentos ou livros raros por dia de trabalho em um único equipamento.

Optou-se, então, pela utilização preferencial de um sistema de digitalização baseado no uso de câmera digital de boa qualidade (com operação relativamente simples) e pelo treinamento e capacitação dos pesquisadores para essa tarefa, aumentando ainda mais a capacidade produtiva das equipes de campo.

Considere-se ainda que, para maior segurança do original, cada página ou face do original foi registrada digitalmente por três vezes. Esse nível de redundância foi utilizado como forma de garantir que, em caso de incidente qualquer, ao menos uma imagem digital estaria em perfeitas condições quando fosse consultada em qualquer tempo futuro.

Tal nível de produtividade, sem um sistema de catalogação adequado, poderia resultar em sério desperdício de esforços. Seria praticamente impossível encontrar uma determinada imagem sem identificá-la corretamente, relacionando-a diretamente ao original.

Foi adotada uma série de critérios para o registro adequado das informações dos documentos originais, para o correto armazenamento dos arquivos digitais de imagens gerados, além de protocolos e procedimentos de conferência da precisão das informações registradas.

Para o armazenamento dos dados e das imagens foi utilizado um servidor com sistema operacional *Linux*, com 2 *terabytes* de espaço de armazenamento e foi utilizada a tecnologia *RAID-5*.⁸

Um sistema de backup das informações captadas foi implantado, garantindo que, no mínimo, duas cópias transitórias dos arquivos digitais gerados pelas equipes de campo estivessem presentes nos discos rígidos dos computadores por eles utilizados. Essas informações foram transferidas para o servidor de dados e de imagens do projeto, utilizado unicamente para esse fim, e foram realizadas cópias de segurança em discos de DVD. Essas cópias foram armazenadas em local seguro e diferente do local do servidor do projeto. Após isso, os arquivos transitórios deveriam ser apagados, dando espaço para um novo lote de imagens digitalizadas. Isso implicou diretamente que não haveria cópias não autorizadas das imagens circulando indevidamente. Um servidor reserva também foi adotado como parte da política de contingência.

O tamanho dos discos de dados utilizados correspondia, então, a uma semana de trabalho contínuo de digitalização, dando a margem de tempo necessária para a execução adequada da transposição e produção de cópias de segurança. Tudo devidamente registrado e comunicado entre as equipes de campo e a administração do projeto.

Sabendo-se que os discos de dados são produzidos em escala industrial e que a produção industrial implica em determinadas margens de tolerância a falhas, aplicadas de forma diferentes pelos diferentes fabricantes, adotou-se as sugestões do *Institute of Electrical and Electronics Engineers (IEEE MSST 2001)*, *Mass Storage Systems and Technologies* como referência de prevenção para a possibilidade de falhas na integridade das informações armazenadas.

A fim de se estender ao máximo a vida útil de toda a informação digital, adotou-se o critério de uso de discos de dados de grande qualidade, entre os melhores fabricados e disponíveis na época da execução do projeto. Estabeleceu-se a sua respectiva troca a cada 5 anos, conforme recomendações dos próprios fabricantes e os estudos das grandes instituições de pesquisa, em nível

⁸ RAID - Redundant Array of Independent Disks ou Matriz redundante de Discos Independentes. É uma tecnologia onde as mesmas informações são armazenados em vários discos (redundância). Se um disco falhar os demais continuarão funcionando e disponibilizarão os dados do disco que falhou. Têm por finalidade, oferecer segurança e confiabilidade dos dados armazenados, evitando perdas.

internacional como as universidades de Cambridge, MIT e o Supercedes Carnegie Mellon University Parallel Data Lab.

Os documentos digitais das imagens armazenados no servidor deram origem a cópias de consulta para visualização em monitores, além de miniaturas para rápida identificação e para publicação na internet (objetivo do projeto). Essas cópias foram produzidas por sub-rotinas automáticas, implementadas no próprio servidor de imagens do projeto. Cópias de segurança foram produzidas e armazenadas em outro computador na própria sede do projeto.

Todo o processo de leitura e de transcrição dos documentos foi feito a partir dessas cópias de consulta, reduzindo consideravelmente o tráfego na rede interna da sede do projeto e requerendo menor poder de processamento nos computadores utilizados nas estações de trabalho. Otimizou-se, assim, o fluxo das atividades relacionadas à pesquisa e consulta de uma obra digitalizada.

O software foi desenvolvido sob medida, de acordo com as necessidades e usos identificados na fase de planejamento. Foi utilizada a linguagem de programação *Hypertext Preprocessor* (PHP) e MySQL para armazenamento dos dados.

A referência universal para todos os tipos de documentos digitais, (imagens ou textos) e sua correta localização nos servidores e discos de dados foi o número de inventário adotado para cadastrar a obra original, que referenciou todas as etapas e processos envolvidos, da identificação do original, passando pela digitalização até a sua publicação na página de internet do projeto.

Em linhas gerais, pode-se dizer que foi uma operação com alto nível de eficiência gerando um mínimo risco à segurança dos originais, custo relativamente baixo por unidade digitalizada e de grande precisão. Essa eficiência foi alcançada devido à adoção de estratégias firmemente calçadas nas mais avançadas técnicas e procedimentos em nível mundial, nas diversas áreas de atuação envolvidas na sua execução, resultando em benefícios de longo prazo maiores do que aqueles que foram traçados na concepção original do projeto.

4 CONCLUSÕES

Este estudo teve como proposta, a partir de constatação de uma situação-problema – a necessidade de salvaguardar um conjunto documental histórico em equipamento cultural público, ou seja, garantir sua conservação e arquivamento, assim como responder à demanda de democratização do acesso, para diferentes tipos de públicos, a bens culturais - investigar soluções propostas pela tecnologia da informação.

Atento às políticas públicas voltadas para estas demandas, assim como ao desenvolvimento da tecnologia da informação, buscou-se no Gerenciamento Eletrônico de Documentos (GED) instrumental a ser explorado, enquanto estratégia de encaminhamento de procedimentos técnico-administrativos.

No entanto, levando-se em conta a exigüidade de tempo para o desenvolvimento de uma análise de viabilidade, treinamento de pessoal, aquisição de equipamentos, elaboração e execução de projeto piloto, optou-se por um estudo da ferramenta GED que incluísse um levantamento de procedimentos, com a identificação de tecnologia e sua aplicabilidade, assim como, análise de projeto já em funcionamento, entendido como “experimentação”, ou seja, situação-laboratório que permitisse confirmar procedimento, assim como avaliar sua eficácia.

A partir das premissas apresentadas, considerou-se importante o conhecimento do processo e das ferramentas do GED para que se pudesse aplicar a tecnologia adequada à solução dos problemas existentes no Museu Paulista. Da mesma forma, observou-se a necessidade de se identificar uma metodologia eficaz, que definisse as etapas a serem percorridas na implantação do Gerenciamento Eletrônico de Documentos. Assim, tomou-se o planejamento como o ponto determinante para o sucesso da empreitada.

O trabalho em conjunto de profissionais de Tecnologia da Informação (TI) e arquivistas representa, no caso, um diferencial quanto à eficiência do sistema a ser implantado. De um lado os profissionais de TI, responsáveis pelo estudo das tecnologias, definição de hardware e software, aspectos ligados à rede de dados, capacidade de armazenamento, processos de cópias de segurança, segurança das informações entre outros aspectos. De outro, os arquivistas, que devem definir os campos e índices necessários à recuperação das informações armazenadas. A

pesquisa sobre procedimentos também identificou como parte das atribuições dos arquivistas o desenvolvimento de um vocabulário controlado, ou seja, palavras-chaves de busca, etapa essencial para a organização e recuperação dos documentos armazenados.

Por meio da atuação conjunta dessas áreas (arquivologia e TI), viabiliza-se o funcionamento de Banco de Dados a ser disponibilizado através de *browsers* ou navegadores, que possibilitam a utilização na Internet, facilitando o acesso ao acervo, como “porta de entrada” à Instituição.

No primeiro capítulo procurou-se entender o que é o GED, onde se aplica e suas tecnologias, visando o conhecimento para aplicação das tecnologias adequadas ao Museu Paulista da USP. No segundo capítulo, apresentou-se a instituição Museu Paulista da USP para que o leitor tome conhecimento da instituição, seus acervos e o processo de informatização já existente, em relação ao acervo. No capítulo 3 foi tomado conhecimento das dificuldades encontradas pela instituição e procurou-se definir um caminho (metodologia) que propiciasse planejar um projeto-piloto de implantação do GED. Neste capítulo também foram apresentadas diretrizes e padrões adotados para os sistemas de GED no intuito de auxiliar no processo de implantação.

Por fim procurou-se tomar conhecimento de um projeto implantado com sucesso, no caso o Projeto José Bonifácio, tendo a finalidade de compreender procedimentos e soluções empregadas, contribuindo com as estratégias a serem adotadas, em relação à informatização do acervo textual do Museu Paulista.

Na atual sociedade da informação, percebe-se que o acesso ao conhecimento é um dos componentes para a diminuição das desigualdades sociais. A tecnologia da informação (TI) apresenta-se como mecanismo eficiente para a produção e compartilhamento das informações. Entre elas, o Gerenciamento Eletrônico de Documentos, através de seu grupo de tecnologias incorporadas, mostra-se uma ferramenta ideal para a formação, organização, recuperação e preservação do acervo documental do Museu Paulista. Ter acesso ao conjunto documental de forma rápida e eficiente sem que haja manipulação desses documentos históricos, além de disponibilizar o acervo ao público em geral e pesquisadores estimulando a pesquisa, consolida a necessidade de implantação do GED no Museu Paulista da USP.

REFERÊNCIAS

BALDAM, R.; VALLE, R.; CAVALCANTI, M. **Ged Gerenciamento Eletrônico de Documentos**. São Paulo: Érika, 2002. 204 p.

BASTOS, Jorge. Projeto José Bonifácio. [Entrevista a Luciano Antonio Beraldo em visita a sua empresa], São Paulo. 15 jan. 2009.

BIBLIOTECA MARIO DE ANDRADE. São Paulo. Disponível em: <http://portal.prefeitura.sp.gov.br/secretarias/cultura/bma/sobre_biblioteca/0001>. Acesso em: 02 fev. 2009.

BOSTELMANN, J. B.; CUNHA, M.A.V.C. da; PAHL, M.T.R. **Desenvolvimento de pacotes específicos versus pacotes de mercado**. Paraná. 1991. Disponível em: <<http://www.pr.gov.br/batebyte/edicoes/1991/bb11/desenvolvimento.htm>>. Acesso em: 05 dez. 2008.

CÂMARA TÉCNICA DE DOCUMENTOS ELETRÔNICOS. **Modelos de requisitos para sistemas informatizados de gestão arquivística de documentos e-ARQ Brasil**. Rio de Janeiro 2006. Disponível em: <<http://www.conarq.arquivonacional.gov.br/Media/publicacoes/earqbrasilv1.pdf>>. Acesso em: 20 jan. 2009.

CENTRO NACIONAL DE GESTÃO DA INFORMAÇÃO. **Aplicabilidades do GED**. São Paulo. Disponível em: <<http://www.cenadem.com.br/ged07.php>>. Acesso em: 12 dez. 2008b.

_____. **As tecnologias correlatas do GED**. São Paulo. Disponível em: <<http://www.cenadem.com.br/ged04.php>>. Acesso em: 15 dez. 2008d.

_____. **Cenadem Trabalhando com você para a divulgação do GED**. São Paulo. Disponível em: <<http://www.cenadem.com.br/empresa.php>>. Acesso em: 10 dez. 2008a.

_____. **Durabilidade das mídias**. São Paulo. Disponível em: <<http://www.cenadem.com.br/ged10.php>>. Acesso em: 16 dez. 2008e.

_____. **ECM e GED são a mesma coisa ?**. São Paulo, 2004. Disponível em: <<http://www.cenadem.com.br/ged03.php>>. Acesso em: 11 dez. 2008.

_____. **Glossário de tecnologia**. São Paulo. Disponível em: <<http://www.cenadem.com.br/ged04quadro.php>>. Acesso em: 10 dez. 2008f.

_____. Por que as empresas no Brasil estão implantando o GED? . Disponível em: <<http://www.cenadem.com.br/ged08.php>>. Acesso em: 10 dez. 2008c.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Guia para a digitalização de documentos versão 2**. Brasília, 2006. Disponível em: <<http://www.sct.embrapa.br/goi/manuais/GuiaDigitalizacao.pdf>>. Acesso em: 15 dez. 2008.

FERREIRA, Maurício A. **Qualidade & Produtividade!!! : Por que utilizar GED ou ECM?** São Paulo. Disponível em: <<http://neeware.com.br/xoops/modules/news/article.php?storyid=3>>. Acesso em: 15 dez. 2008.

FUNDAÇÃO BIBLIOTECA NACIONAL. Rio de Janeiro. Disponível em: <<http://www.bn.br/portal/>>. Acesso em: 02 fev. 2009.

HADDAD, Samir R. **GED – uma alternativa viável na gestão da informação estratégica**. 2000. 39 p. Monografia (Curso de Especialização em Informática Pública) – Universidade Católica de Minas Gerais, Minas Gerais, 2004. Disponível em: <<http://www.powerbrasil.com.br/pdf/haddad2000.pdf>>. Acesso em: 18 dez. 2008.

MAMELUCO. São Paulo. Disponível em: <<http://www.mameluco.com.br/>>. Acesso em: 02 fev. 2009.

MUSEU HISTÓRICO NACIONAL. Rio de Janeiro. Disponível em: <<http://www.museuhistoriconacional.com.br/>>. Acesso em: 02 fev. 2009.

MUSEU DO ÍNDIO. Rio de Janeiro. Disponível em: <<http://www.museudoindio.org.br/>>. Acesso em: 02 fev. 2009.

MUSEU MINEIRO. Minas Gerais. Disponível em: <<http://www.cultura.mg.gov.br/?task=interna&sec=3&cat=45&con=377>>. Acesso em: 02 fev. 2009.

SAMARA, Eni de Mesquita. Museu Republicano preserva história. São Paulo. Disponível em: <<http://www.usp.br/mapas/materia03.html>>, Acesso em: 10 dez. 2008.

SILVA, Antônio P. de A e. **ECM/GED: tecnologia para tratar documentos, informações e conteúdo**. São Paulo. Disponível em : <http://www.arquivar.com.br/espaco_profissional/noticias/mercado-tecnologia/ecm-ged-tecnologia-para-tratar-documentos-informacoes-e-conteudo>. Acesso em: 03 fev. 2009.

STRINGHER, Ademar. **Aspectos Legais da Documentação em Meios Micrográficos, Digitais e Eletrônicos**. São Paulo: Cenadem, 2003. 242p.

TAKAHASHI, Tadao. **Sociedade da informação no Brasil: livro verde**. Brasília: Ministério da Ciência e Tecnologia, 2000. 140p.

THE ASSOCIATION FOR INFORMATION AND IMAGE MANAGEMENT. **About AIIM**. EUA. Disponível em: <<http://www.aiim.org/AboutAIIM/ECM-ERM-BPM-Association.aspx>>. Acesso em: 20 jan. 2009.

THE ASSOCIATION FOR INFORMATION AND IMAGE MANAGEMENT. **Como montar projetos de GED gerenciamento eletrônico de documentos, linhas mestras para análise, seleção e implantação**. Tradução de Ana Claudia Schmidt. São Paulo: Ed. Cenadem, 2008. 90p.

THE ASSOCIATION FOR INFORMATION AND IMAGE MANAGEMENT. **What is ECM**. USA. Disponível em: <<http://www.aiim.org/ResourceCenter/AboutECM.aspx>>. Acesso em: 12 dez. 2008.

WILLIS, Don. Uma abordagem de sistemas híbridos para a preservação de materiais impressos. Rio de Janeiro, 2001. Disponível em: <http://143.106.151.46/cpba/pdf_cadtec/50.pdf>. Acesso em: 16 jan. 2009.

WITTER, José Sebastião. et al. **Acervos do Museu Paulista/USP (Museu do Ipiranga)**. São Paulo: Imprensa Oficial do Estado, 1999. 127p.

WORKFLOW MANAGEMENT COALITION. USA. Disponível em: <<http://www.wfmc.org/>>. Acesso em: 20 fev. 2009.

ZEQUINI, Anicleide. **Acervo Arquivístico do Museu Republicano "Convenção de Itu"** Itu: Museu Paulista - USP. 2000. Disponível em: <<http://www.geocities.com/arquivosmunicipais/itu.htm>>. Acesso em: 20 dez. 2008.

REFERÊNCIAS CONSULTADAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR10520**: Informação e documentação: citações em documentos – Apresentação. Rio de Janeiro, 2002.

_____. **NBR 6023**: Informação e documentação: referência – elaboração. Rio de Janeiro, 2002.

BRYANT, Duhon; JEETU, Patel; TUCKER, Rick. **ECM na sua empresa**. Tradução de Ana Claudia Schmidt. São Paulo: Cenadem, 2005/2006. 34 p.

MACEDO, Geraldo Majela Ferreira de, **Bases para a implantação de um sistema de gerenciamento eletrônico de documentos – GED**. 2003. 144 p. Dissertação (Mestrado) – Universidade Federal de Santa Catarina. Disponível em: <<http://www.tede.ufsc.br/teses/PEPS2944.pdf>>. Acesso em: 11 dez. 2008.

REIS, M.G.C.; et.al. **Diretrizes para apresentação de dissertações e teses**. 3.ed. São Paulo : Divisão de Biblioteca da EP/USP, 2006. 103 p.

THE ASSOCIATION FOR INFORMATION AND IMAGE MANAGEMENT.
Recommended Practice Analysis, Selection, and Implementation of Electronic Document Management Systems (EDMS). USA. 2007. 47 p. Disponível em: <<http://www.aiim.org/documents/standards/arp1-2007.pdf>>. Acesso em: 10 fev. 2009

VALLE, Eduardo. **Sistemas de informação multimídia na preservação de acervos permanentes**. 2003. 128 p. Dissertação (Mestrado) – Universidade Federal de Minas Gerais. Disponível em: <<http://eduardovalle.com/docs/valle03dissertacaoPreservacaoArcervos.pdf>>. Acesso em: 15 dez. 2008.

ANEXO A – Glossário das Tecnologias

CAPTAÇÃO É o modo como as informações e documentos, eletrônicos ou em papel, passam para um repositório de conteúdo para reutilizar, distribuir e armazenar.	ARMAZENAMENTO É onde se coloca o conteúdo e como você o encontra novamente.	GERENCIAMENTO São as ferramentas e técnicas para mover conteúdo pela organização e monitorar o desempenho delas.	DISTRIBUIÇÃO Como você consegue o conteúdo certo para o público certo no dispositivo certo. (Acesso).	PRESERVAÇÃO As suas opções para armazenamento a longo-prazo do conteúdo essencial de sua empresa.
<p>Agregação O processo de combinar a entrada de dados de criação e autoria de diferentes ferramentas e outros sistemas.</p> <p>COLD/ERM Armazena e indexa saídas de computador (principalmente relatórios) em discos magnéticos, discos ópticos e fitas magnéticas. Uma vez armazenados, os relatórios podem ser acessados, visualizados, impressos, transmitidos por fax ou distribuídos pela Internet. Frequentemente usado para aplicações de contas pela Internet.</p> <p>Categorização Organiza documentos, páginas web e outros conteúdos em um agrupamento lógico baseado no próprio conteúdo.</p> <p>Document Imaging Processo de captação, armazenamento e busca de documentos, independentemente do formato original, usando microformas e/ou imagem digital (digitalização, OCR, ICR etc).</p> <p>Eforms/WebForms Formulários desenhados, gerenciados e completamente processados num ambiente eletrônico.</p>	<p>Trilhas de auditoria Registro de quem fez mudanças, quando e quais, para efeito de contabilidade.</p> <p>CD-ROM Compact Disc Read Only Memory – Disco óptico criado por um processo de masterização e usado para distribuir informação apenas de leitura.</p> <p>Check In-Check Out Assegura que somente uma pessoa pode trabalhar num documento a qualquer tempo.</p> <p>Content Management System A capacidade de gerenciar e rastrear o local e relacionamento entre conteúdos dentro de um repositório.</p> <p>Data Warehouse Repositório central para todos, ou a maioria, dos dados estruturados da organização.</p> <p>Database 1 – Coleção eletrônica de registros armazenados num arquivo central e acessível por muitos usuários e para muitas aplicações. 2 – Coleção de elementos de dados dentro de registros ou arquivos que têm relacionamento com outros registros ou arquivos. Banco de dados relacionais são mais comuns – o dado é armazenado em linhas, colunas e tabelas padronizadas. Banco de dados em XML são tecnologia em desenvolvimento.</p>	<p>Colaboração Ferramentas (autoria colaborativa, videoconferência, quadros compartilhados etc) que permitem a usuários trabalharem o mesmo conteúdo num ambiente comum.</p> <p>Document Management – Software que controla e organiza documentos pela empresa, incorpora captação de conteúdo e documento, Workflow, repositórios de documento, sistemas de saída e COLD/ERM e sistemas de busca de informação.</p> <p>Web Content Management Uma tecnologia que direciona a criação do conteúdo, revisão, aprovação e processo de publicação de conteúdo baseado na web.</p> <p>Workflow/BPM (Gerenciamento de Processo de Negócio) Automação de processos de negócio no todo ou em parte, em que documentos, informação ou tarefas são passadas de um participante a outro para ação, de acordo com regras estabelecidas. Um processo de negócio é um conjunto de fluxos de trabalho logicamente relacionado, passos de trabalho e tarefas que provêm um produto ou serviço para clientes. BPM é um mix de gerenciamento de processos e Workflow com tecnologia de integração de aplicação.</p>	<p>COLD/ERM Saída de Computador para Disco Laser/Gerenciamento Eletrônico de Relatórios.</p> <p>Compressão Técnica usada para reduzir o número de bits num arquivo de imagem digital, como Tiff e JPEG.</p> <p>Gerenciamento de Direitos Digitais Permite a distribuição segura e desabilita distribuição ilegal de conteúdo pago na web.</p> <p>Assinatura Digital Usada para autenticar o remetente da mensagem.</p> <p>PDF - Portable Document Format Formato desenvolvido pela Adobe Systems para publicação de documentos.</p> <p>Personalização Relaciona o conteúdo com o indivíduo.</p> <p>PKI Habilita a troca segura de conteúdo através do uso de um par de chaves pública e privada criptografada, obtida por autoridade confiável.</p> <p>Syndication (Associação) Provisão de conteúdo para reutilizar e integração com outros materiais, geralmente através de pagamento de uma assinatura.</p>	<p>Microfilme (Cartão-janela, microficha, jaquetas, rolo de 16mm) – 1 - Grão fino de filme de alta resolução usado para registrar imagens reduzidas em tamanho em relação ao original. 2- Microforma no formato de tira ou rolo. O registro micrográfico num filme.</p> <p>Disco óptico Originalmente, WORM – Write-Once, Read Many. Disco óptico no qual dados são gravados pelo usuário uma vez (e é inalterável) e pode ser lido muitas vezes.</p> <p>Papel Secular e, como o microfilme, um dos dois modos de assegurar que documentos sejam lidos daqui a cem anos ou mais.</p> <p>SAN, NAS CAS Todos estão sendo progressivamente usados para armazenar conteúdo. CAS é uma metodologia de arquivamento planejada para rápido acesso a conteúdo fixo.</p>

continua

continuação

CAPTAÇÃO É o modo como as informações e documentos, eletrônicos ou em papel, passam para um repositório de conteúdo para reutilizar, distribuir e armazenar.	ARMAZENAMENTO É onde se coloca o conteúdo e como você o encontra novamente.	GERENCIAMENTO São as ferramentas e técnicas para mover conteúdo pela organização e monitorar o desempenho delas.	DISTRIBUIÇÃO Como você consegue o conteúdo certo para o público certo no dispositivo certo. (Acesso).	PRESERVAÇÃO As suas opções para armazenamento a longo-prazo do conteúdo essencial de sua empresa.
<p>Forms Processing A habilidade de um software aceitar formulário digitalizado e extrair dados das caixas (campos) e linhas para alimentar banco de dados. Software usualmente inclui a habilidade de tirar a máscara do formulário para aumentar o nível de reconhecimento. Reconhecimento inteligente de documento identifica automaticamente os tipos de documentos através de layout e estrutura do documento.</p> <p>HCR Reconhecimento de Caracter Manuscrito – Tecnologia OCR desenvolvida para transformar imagens de caracteres manuscritos em código ASCII.</p> <p>ICR Reconhecimento Inteligente de Caracteres – Forma avançada de tecnologia OCR que pode incluir capacidades como aprender fontes durante o processo ou usar o contexto para fortalecer probabilidades de reconhecimento correto ou reconhecer caracteres manuscritos.</p> <p>OCR Reconhecimento Óptico de Caracteres – Técnica pela qual imagens de caracteres podem ser identificadas por máquina e assim serem convertidas em códigos processáveis por computadores.</p>	<p>DVD Digital Versatile Disc – Disco óptico de 120 mm no qual vídeo digital, áudio, dado e imagens podem ser armazenadas. Disponível nos formatos R, RW, e ROM.</p> <p>Sistema de Arquivo(Diretórios) O modo pelo qual arquivos são nomeados e onde estão alocados logicamente para armazenamento e busca mais comumente numa estrutura hierárquica de árvore (diretórios).</p> <p>Magneto-óptico Gravar dados usando combinação de meios ópticos e magnéticos para trocar a polaridade de um campo magnético na mídia de gravação. O dado pode ser apagado ou regravado.</p> <p>NAS Pode ser parte da SAN. Armazenamento direto no HD ligado a rede de trabalho para prover acesso à informação.</p> <p>Disco óptico Mídia que aceita e retem informação na forma de marcas ou densidade de modulação numa camada de gravação que pode ser lida com um feixe de luz óptico.</p> <p>RAID Redundant Array of Independent Disks – Armazena os mesmos dados em múltiplos HDs para melhorar o desempenho e tolerância a falhas.</p>	<p>Gestão Documental - Records Management Capacita uma empresa a atribuir um ciclo de vida específico a uma peça individual de informação corporativa da criação, recebimento, manutenção e uso até a última disponibilização dos registros. Um registro não é, necessariamente, o mesmo que um documento. Todos os documentos são potenciais registros, mas não vice-versa. Um registro é essencial para o negócio, documentos são containers de informação para se trabalhar. Registros são documentos com valor de evidência.</p>	<p>Transformação A troca de formato de um conteúdo para outro, conforme a necessidade para a distribuição.</p> <p>XML Um padrão estabelecido, baseado na SGML – Standard Generalized Markup Language, desenvolvida para facilitar a construção de itens padronizados de dados. Também usado como um mecanismo genérico de troca de dados.</p>	

continua

continuação

CAPTAÇÃO É o modo como as informações e documentos, eletrônicos ou em papel, passam para um repositório de conteúdo para reutilizar, distribuir e armazenar.	ARMAZENAMENTO É onde se coloca o conteúdo e como você o encontra novamente.	GERENCIAMENTO São as ferramentas e técnicas para mover conteúdo pela organização e monitorar o desempenho delas.	DISTRIBUIÇÃO Como você consegue o conteúdo certo para o público certo no dispositivo certo. (Acesso).	PRESERVAÇÃO As suas opções para armazenamento a longo-prazo do conteúdo essencial de sua empresa.
<p>OMR Reconhecimento Óptico de Marca – Detecta presença ou ausência de marcas em áreas definidas, usado para processar questionários, testes padronizados etc.</p> <p>Indexação Identificação de atributos específicos de um documento ou registro da base de dados para facilitar a busca.</p> <p>Design de entrada Modelos usados para habilitar autores para facilmente colocar conteúdo num sistema customizado, baseado no tipo e formato do conteúdo que será inserido.</p> <p>XML eXtensible Markup Language – Um padrão estabelecido, baseado na SGML – Standard Generalized Markup Language, desenvolvida para facilitar a construção de itens padronizados de dados. Também usado como um mecanismo genérico de troca de dados.</p>	<p>Repositórios Parte do sistema de Document Management. Funcionalidade específica para controlar o check in/out de material, controle de versão e pesquisa junto a atributos definidos.</p> <p>Retrieval Procedimento de busca e de extração de registros e conteúdo da base de dados.</p> <p>SAN Uma rede de trabalho com alta velocidade que conecta sistemas de computadores e elementos de armazenamento e permite o movimento de dados entre o sistema de computadores e elementos de armazenamento e entre os elementos de armazenamento.</p> <p>Fita Mídia de armazenamento magnética. Tamanhos padrão são 8mm, 1/8 polegada, 1/4 polegada, 1/2 polegada, 4mm DAT (Digital Audio Tape) e DTL (Digital Linear Tape) em rolos ou cassetes.</p> <p>Controle de versão Procedimentos para identificar o autor e as seqüências de diferentes versões do documento.</p>			

Fonte: (Cenadem, 2008f).

O conceito definido pela AIIM é de *Enterprise Content Management* (ECM) e não de GED. Para a compreensão desse conceito, deve-se entender primeiramente a diferença entre as nomenclaturas. A partir do desenvolvimento das tecnologias em diversos países (norte-americanos ou europeus), quando chegaram ao Brasil, apesar de trazerem nomenclaturas diferentes, eram assemelhadas. Daí, certa indefinição e, até mesmo, sobreposição de tecnologias e nomenclaturas.

1.3 SIGLAS

Atualmente, é isso que está acontecendo conceitualmente, por exemplo, com o ECM e o *Integrated Document System* (IDM) que, correspondem para nós técnicos brasileiros, ao GED e ao *Document Management* (DM) (SILVA, 2009).

O ECM fala de conteúdo (todas as informações que constituem conhecimento para a organização em formato analógico ou digital), e todo esse conteúdo precisa ser administrado. As tecnologias da informação forneceram produtos para administrar toda essa massa de conteúdo e ganharam vários nomes, entre eles o ECM. Independente do conteúdo, sistemas de GED sempre contemplaram o gerenciamento da informação, ou seja: o GED sempre administrou conteúdos (ibidem).

O GED engloba as tecnologias de *Document Imaging* (DI), *Document Management* (DM), *Records Management* (RM), *Computer Output to Laser Disk* (Cold), *Enterprise Report Management* (ERM), *Workflow* e outras. O ECM estende as tecnologias cobertas pelo GED, incluindo: *Web Content Management* (WCM), *Digital Asset Management* (DAM), *Digital Rights Management* (DRM) e outras (FERREIRA, 2008).

Podemos então concluir que a principal diferença entre GED e o ECM está no leque de tecnologias oferecidas por cada uma delas, as quais serão apresentadas adiante. Tendo em vista a semelhança entre elas, consideraremos neste trabalho o GED e o ECM equivalentes.

A partir das considerações expostas, será apresentado o conceito de ECM definido pela AIIM: São as tecnologias, ferramentas e métodos usados para captar, gerenciar, armazenar, preservar e distribuir conteúdo pela empresa. No nível mais

elementar, as ferramentas e estratégias de ECM permitem o gerenciamento de informação não-estruturada de uma organização, enquanto aquela informação existir (AIIM, 2008).

A partir das definições das três entidades, é possível determinar alguns princípios básicos do Gerenciamento Eletrônico de Documentos. O GED é um conjunto de tecnologias, que utiliza computadores, para a organização da informação existente em documentos e dados eletrônicos. Chama-se sistema de gerenciamento eletrônico porque as mídias usadas para guardar os documentos são eletrônicas, como discos magnéticos e ópticos.

Essas tecnologias podem ser usadas isoladamente ou em conjunto. Foram divididas em cinco funções básicas para o gerenciamento das informações: captura, gerenciamento, armazenamento, distribuição e preservação. O GED deve possuir um modo de visualização dos documentos, simplesmente o gerenciamento das informações contidas nos documentos não pode ser considerado um sistema de GED. A informação deverá ser gerenciada enquanto ela existir, ou seja, em todo seu ciclo de vida.

Entende-se como o ciclo de vida fases que o documento passa enquanto de sua existência, entre elas: a produção, tramitação, uso, avaliação, arquivamento e eliminação. Em seu ciclo de vida os documentos passam por três idades: a corrente, no qual os documentos são consultados freqüentemente, estando arquivados ou não; a intermediária onde os documentos possuem pouca utilização e aguardam o tempo, para eliminação ou guarda no arquivo permanente; a permanente onde se guardam definitivamente os documentos em razão de seu valor legal ou histórico (CONARQ, 2006).

1.4 APLICABILIDADES DO GED

O Cenadem apresenta uma série de aplicabilidades para o GED. Como poderá ser visto abaixo, a conversão de acervos históricos inclui-se entre suas aplicabilidades (CENADEM, 2008b).

- Apoio aos processos de fiscalização;
- Apoio documental aos sistemas de *Geographic Information System* (GIS);

- Arquivos de recortes de jornais e revistas - *Clipping*;
- Atendimento a clientes de serviços utilitários: telefonia, energia elétrica e outros;
- Atendimento aos clientes de banco: extratos de conta corrente, aplicações;
- Automação de cartórios;
- Bibliotecas digitais;
- Cartões de assinatura;
- Catálogos de peças e listas de preços;
- Contratos de câmbio;
- Contratos de financiamento e leasing;
- Controle completo de compras. Do pedido à entrega do produto;
- Controle de bilhetes de companhias aéreas;
- Controles de documentos de arrecadação: impostos, taxas e multas em organismos do governo;
- **Conversão de acervos históricos;**
- Conversão de sistemas micrográficos;
- Depósitos e pagamentos na retaguarda das agências bancárias;
- Desenhos de engenharia e relatórios técnicos;
- Disponibilização ampla de documentos oficiais. Por exemplo, diários oficiais;
- Documentação administrativa de hospitais;
- Documentação cadastral e societária de clientes;
- Documentação da logística de transporte;
- Documentação de auditoria;
- Documentação de consórcios;
- Documentação dos sistemas de qualidade - ISO 9000 por exemplo;
- Documentação e acompanhamento do ciclo de vida do produto;
- Documentação e relatórios contábeis e financeiros;
- Documentos de escritório de modo geral: *Word, Excel, Power Point* e outros;
- Documentos e processos de tribunais;
- Documentos em geral das polícias civis, militares, detrans e outros;
- Documentos em geral de instituições de ensino;
- *Forms Processing* - Processamento de Formulários;

- Gerenciamento da documentação de benefícios. Fundos de Pensão;
- Gerenciamento de contratos em geral;
- Gerenciamento de correspondência, fax, e-mail e outros veículos;
- Gerenciamento de ordens de serviço;
- Gerenciamento de processos de concessão;
- Processamento de cheques. Da consulta à compensação;
- Processos de crédito imobiliário;
- Processos de importação e exportação;
- Prontuários das áreas de recursos humanos, incluindo recrutamento e seleção;
- Prontuários de pacientes em hospitais;
- Resultados de exames laboratoriais;
- Seguradoras: apólice, sinistro até a indenização;
- Apoio ao gerenciamento do conhecimento, *Enterprise Resource Planning* (ERP), *Customer Relationship Management* (CRM), Comércio eletrônico e outras tecnologias.

Os ERPs, chamados de sistemas de gestão integrados, têm por finalidade a integração da informação de vários setores da empresa, com o objetivo de evitar retrabalhos, tornando disponível a informação, como por exemplo: sistema de finanças, contabilidade, compras, vendas, etc. As informações não estruturadas não fazem parte de sistemas ERP, como por exemplo: cheques, contratos, notas fiscais, etc.

O CRM é um sistema que usa ferramentas para conhecer hábitos e necessidades dos clientes. Os documentos fornecidos pelos clientes devem alimentar a base de dados. Exemplo: documentos de atendimentos aos clientes, solicitações, reclamações, sugestões, etc.

Não é intenção deste trabalho, relatar todas as tecnologias de informação, às quais o GED agrega valor. O importante é compreender que o GED participará como ferramenta de auxílio no gerenciamento dessas informações, oferecendo o suporte documental necessário.

1.5 IMPLANTAÇÃO DO GED NO BRASIL

A opção por implantação de GED, em equipamentos culturais e empresas, no Brasil, vem crescendo. Segundo pesquisas realizadas pelo Cenadem, os principais motivos que levam as empresas Brasileiras a implantarem o GED, são (CENADEM, acesso em: 2008c):

- Permite a recuperação da empresa após desastres, como o caso do *World Trade Center*, EUA, em setembro de 2001;
- Redução de custos com cópias, já que há disponibilização dos documentos em rede;
- Absoluto controle no processo de negócio;
- Alta velocidade e precisão na localização de documentos;
- Criação de facilidades para o trabalhador do conhecimento;
- Disponibilização instantânea de documentos sem limites físicos;
- Eliminação de fraudes, principalmente em agências governamentais;
- Gerenciamento automatizado de processos, minimizando recursos humanos e aumentando a produtividade;
- Grande melhoria no processo de tomada de decisões;
- Aproveitamento de espaço físico;
- Ilimitadas possibilidades para indexação e localização de documentos;
- Evitar extravio ou falsificação de documentos;
- Integração com outros sistemas e tecnologias;
- Aproveitamento da base de informática já instalada na sua empresa;
- Mais agilidade nas transações entre empresas;
- Maior velocidade na implementação de mudanças nos processos;
- Exigência da alta competitividade;
- Melhor atendimento ao cliente. O GED proporciona respostas precisas e instantâneas;
- Obtenção de vantagem competitiva sustentável;
- Possibilidade da empresa virtual sem limites físicos, com redução de despesas e permitindo que funcionários trabalhem a partir de casa;

- Única solução para cumprir prazos de recursos em processos, como no caso de multas de trânsito;
- Tecnologia complementar de outras como ERP e CRM.

1.6 TECNOLOGIAS QUE COMPÕEM O GED

O GED não é uma tecnologia, mas sim um conjunto de tecnologias para solucionar os problemas de gerenciamento de documentos em forma digital. Cada tecnologia tem um papel específico dentro das cinco funcionalidades ou grupos que compõem o GED: captura, gerenciamento, armazenamento, distribuição e preservação. A utilização de uma ou outra tecnologia dependerá da atividade que se deseja implementar.

Para que o Museu Paulista tenha subsídios para a escolha da tecnologia adequada, serão apresentadas, dentro de seus grupos, as tecnologias mais comumente encontradas no mercado.

1.6.1 Processo de Captura

O processo de captura como um todo, é bem mais amplo do que apenas transcrever as informações do meio analógico para o meio digital. Ele engloba as seguintes fases:

- Preparação dos documentos;
- Captura e Digitalização dos documentos;
- Indexação.

a) Preparação dos documentos a serem digitalizados

Nesta fase está incluído, a localização e o transporte dos documentos, montagem de lotes, remoção de grampos, reparo dos documentos (se necessário),

entre outros. Por ser um processo executado de forma manual, e se tratando de documentos históricos há uma série de procedimentos a serem adotados no ambiente de digitalização, entre eles: a utilização de luvas e máscaras faciais, não haver alimentos ou bebidas no local, não empilhar os originais, as janelas devem estar fechadas para evitar a entrada de poeira, etc.

b) Captura e Digitalização dos documentos

Capturar um documento significa incorporá-lo a um repositório de conteúdo. Nos sistemas analógicos essa captura se dá com a transcrição das informações contidas em papel ou microfilme para o meio eletrônico, conhecido como digitalização. Esse processo se dá através da utilização de dispositivos de captura como scanner ou máquina fotográfica digital.

Também é possível a utilização de recursos de *Image Enhancement*, que dispõe de mecanismos como: *Despeckling* (remoção de ruídos), *Deskewing* (endireitar imagens tortas), *Edge Detection* (remoção de bordas), etc. Este processo poderá ser utilizado para a melhoria da imagem e remoção de espaços inadequados ou em branco, procurando reduzir o tamanho da imagem e conseqüentemente o tráfego na rede. Após o processo de conversão, é feito o controle de qualidade das imagens e então efetuada nova digitalização, se necessária (HADDAD, 2000).

Terminada essa etapa, o documento deverá retornar a seu local de origem, tratando-se de documentos históricos, os mesmos não poderão ser eliminados. Segundo a lei nº 5.433, de 08 de maio de 1968, em seu artigo 2º, "Os documentos de valor histórico não deverão ser eliminados, podendo ser arquivados em local diverso da repartição detentora dos mesmos" (STRINGLER, 2003, p. 207).

Dispensa-se o processo de captura para os documentos já criados em formato digital como editores de texto, planilhas, etc.

c) Indexação

A indexação proporciona uma forma de localizar, agrupar, recuperar e gerenciar os documentos (HADDAD, 2000). É o processo de inclusão no banco de dados de informações significativas retiradas do documento com a finalidade de localização. A indexação é um dos componentes mais importantes do GED. Se os

documentos não estiverem adequadamente indexados não poderão ser mais recuperados.

1.6.1.1 Tecnologias relacionadas à Captura

a) *Document Imaging* (DI)

O DI é a tecnologia responsável pela captura, arquivamento e recuperação de documentos. Através do processo de captura, o DI é responsável pela conversão de um documento em meio físico (em papel ou microfilme) para o formato digital, através da utilização de dispositivos de digitalização.

Nos sistemas de DI é necessário um programa que permita o gerenciamento das informações contidas no banco de dados, que tem por fim o armazenamento das informações necessárias, inclusive os índices para que o documento possa ser localizado.

Embora haja uma grande variedade de *softwares* de *Document Imaging* no mercado, algumas características são comuns à maioria deles, entre elas:

Mecanismos para o cadastramento das informações e índices; interface de consultas que apresentem o resultado da pesquisa e visualização das imagens dos documentos, possibilidade de emissão de relatórios diversos, além de outras características.

O foco principal do DI é o gerenciamento de documentos estáticos, documentos estes já concluídos e que não sofrerão mais alterações em seu conteúdo. O DI é uma solução ideal para o gerenciamento de documentos em papel.

Os documentos em papel, quando digitalizados, geram-se documentos em formato de imagem. Nesse formato, os documentos não poderão ser localizados através de busca de palavras contidas no próprio texto. Para que essas palavras possam ser localizadas é necessária a conversão desses documentos em formato de imagem para o formato de texto, através das tecnologias como o *Optical Character Recognition* (OCR) ou *Intelligent Character Recognition* (ICR). A eficiência desse sistema dependerá do estado de conservação do documento e a

porcentagem de texto reconhecida após a aplicação dessas tecnologias. A utilização do DI proporciona diversas vantagens em sua utilização, entre elas: velocidade na recuperação e aspectos ligados a preservação e a segurança física dos documentos, além da redução no espaço de armazenamento, entre outros.

b) *Forms Processing*

São tecnologias que permitem reconhecer e extrair as informações dos formulários, normalmente produzidos com esse fim. Através da utilização de tecnologias como o OCR e ICR os dados serão extraídos e alimentarão automaticamente os campos contidos no banco de dados.

O objetivo dessa tecnologia é diminuir os custos referentes à digitação e tornar ágil o processo. Essa tecnologia pode ser aplicada em formulários bancários, pesquisas de modo geral, processamento de gabaritos de prova, entre outros.

c) *Image Enable*

A maioria dos sistemas corporativos trabalha com informações contidas em suas bases de dados. Contudo, às vezes é necessário visualizar os documentos pertencentes ao processo para a tomada de decisões. O *Image Enable* possibilita que diversos sistemas computacionais existentes nas empresas visualizem essas imagens, mesmo em ambientes que não são de Gerenciamento Eletrônico de Documentos.

Essa visualização poderá ser efetuada através de ligações (*links*) diretas dos dados da aplicação com a imagem, ou através de integrações com um ambiente de *Document Imaging* ou *Document Management*. Nessa segunda forma, as imagens poderão ser visualizadas através da execução do próprio aplicativo de GED. Desta forma todos os documentos envolvidos no processo poderão ser visualizados, não só os referenciados. Essa integração também poderá ser feita com qualquer tipo de conteúdo armazenado, como por exemplo: textos eletrônicos e planilhas. Surge então o conceito de *Content Enabled*. Esse tipo de aplicação poderá ser utilizado, por exemplo: para a exibição de notas fiscais referenciadas em sistemas de compras ou contabilidade; pedido ou reclamação do cliente em sistemas de CRM; radiografias de um cliente num prontuário médico, entre outras (BALDAM, 2002).

d) Tecnologias agregadas a Captura

Optical Character Recognition (OCR):

A partir da digitalização de um documento, é produzida uma imagem em formato digital. A partir da aplicação da tecnologia como o *Optical Character Recognition (OCR)*, pretende-se que haja o reconhecimento dos caracteres e que seja gerado um arquivo em formato texto. A principal vantagem do processo consiste na possibilidade de edição do texto, busca de palavras e grande redução no tamanho do arquivo gerado. O OCR é utilizado no reconhecimento de caracteres gerados de forma mecânica, como por exemplo: documentos datilografados ou impressos.

Intelligent Character Recognition (ICR):

Forma avançada da tecnologia de OCR utilizada para reconhecer caracteres manuscritos, inclui capacidades de como apreender fontes ou utilizar o contexto para fortalecer probabilidades de reconhecimento correto.

Optical Mark Recognition (OMR):

Permite o reconhecimento de marcas em áreas definidas, por exemplo: uma prova de vestibular com alternativas.

Barcode (Código de barras):

Permite converter informações contidas em um código de barras para o formato de texto.

Full Text Retrieval (FTR):

São ferramentas que possuem recursos para recuperar documentos a partir de qualquer palavra contida no texto, e não só pelo seu índice. Nesse processo, os

documentos digitalizados deverão passar por um processo de extração de seu conteúdo através da tecnologia de OCR ou ICR.

1.6.2 Gerenciamento

Para as funcionalidades do gerenciamento, são utilizadas técnicas e ferramentas que gerenciem todo o ciclo de vida do documento. Os documentos podem passar por alterações e revisões durante seu ciclo de vida.

1.6.2.1 Tecnologias Relacionadas ao Gerenciamento

a) *Document Management (DM)*

O DM serve para controlar documentos desde a sua criação até seu descarte, ou seja controla a vida do documento. Seu objetivo é controlar todo documento que esteja em atividade na empresa, documento este em papel ou em formato digital, arquivado ou não (BALDAM, 2002).

Aplica-se ao gerenciamento de documentos dinâmicos, documentos estes que sofrem constantes processos de revisão, como por exemplo: documentos oriundos de editores de texto, planilhas, arquivos de projetos, apresentação de slides, entre outros. O DM através do processo de indexação, controla o acesso a essa massa documental.

Hoje em dia a maior parte dos documentos é gerada de forma digital, documentos estes que sofrem alterações constantes. Um dos grandes problemas que acontecem nas empresas é a gravação do mesmo arquivo várias vezes com nomes e/ou versões diferentes. O DM, através do controle de versões, possibilita a visualização de versões mais antigas, porém permite apenas a utilização da versão mais recente, a criação de novas versões mediante solicitação, o estabelecimento de regras de segurança para acesso a versões anteriores (BALDAM, 2002).

Outra função típica de um ambiente de DM é permitir que se criem diversos modelos de formulários (*templates*) a serem disponibilizados aos usuários com a finalidade de geração de documentos padronizados. Também permite que a partir de dados armazenados num ambiente de DI, sejam automaticamente disponibilizados a um *template* num editor de texto, por exemplo, além de outras funções (BALDAM, 2002). Seu principal foco está no controle das versões, nas datas das alterações efetuadas pelos usuários e no histórico da vida do documento (Cenadem, 2008d).

b) *Engineering Document Management System (EDMS)*

Um sistema de EDMS é, na essência, um sistema de DM, mas, com características específicas para que possa manipular documentos técnicos como: plantas, desenhos, manuais, especificações técnicas, etc. Essa categoria de documentos exige um tratamento especial devido a características próprias, como por exemplo: o tamanho físico, pois aplicações de DM normalmente não suportam visualizações e impressões de arquivos com grandes dimensões; outro aspecto refere-se ao tempo de vida de um documento técnico, que geralmente está relacionado com a vida do equipamento, instalação ou processo que está relacionado, entre outros (BALDAM, 2002).

c) *Records Management (RM)*

A tecnologia de RM destina-se ao gerenciamento da guarda dos documentos. Através dela faz-se o controle da movimentação dos mesmos e gerencia-se o tempo de vida dos documentos. Aplicando-se uma tabela de temporabilidade documental, baseada em aspectos legais ou regulamentações internas, é possível determinar por quanto tempo os documentos deverão ser mantidos. Sua principal justificativa de utilização é a redução dos custos e espaços de armazenagem através do descarte dos documentos desnecessários.

d) *Workflow*

O *Workflow*, do inglês Fluxo de Trabalho, pode ser entendido como uma

ferramenta para a automatização de processos, bem como fazer a análise, compreensão e monitoramento das atividades de negócio.

Um processo de negócio consiste num conjunto de atividades coletivas e relacionadas que visam atingir um determinado objetivo. Segundo a *Workflow Management Coalition* (WFMC), *Workflow* pode ser entendido como uma automação do processo de negócio, no todo ou em parte, onde tarefas, informações ou documentos são passados de um participante para outro, através de uma seqüência de atividades e regras que estabelecem a forma com que se deve dar andamento aos processos (WFMC, 2009).

O *Workflow* é um processo dinâmico, e faz com que a informação percorra a trajetória prevista. Possibilita a coordenação das atividades executadas, como por exemplo: quem são os usuários envolvidos, quais as informações geradas por eles, qual o prazo limite para execução de uma determinada tarefa, entre outras.

Sua utilização assegura que as atividades previstas sejam realizadas assim que possível, por uma pessoa previamente estabelecida e numa seqüência pré-definida, tal como o processo de atendimento a reclamações, compras, etc.

As justificativas para implementação são: redução de custos, aumento da produtividade, controle do trâmite dos processos, eliminação da tramitação física de documentos, entre outros.

1.6.3 Armazenamento

Tem por finalidade armazenar as informações em formato digital para posterior recuperação. Os documentos podem ser armazenados em diferentes mídias:

Ópticas

- *Compact Disc Read Only Memory* (CD-ROM)
Capacidade de armazenamento: 700 Mb
- *Digital Versatile Disk* (DVD)
Capacidade de armazenamento: 4.7 GB (camada única)
8.5 GB (camada dupla)

- *High Density Digital Versatile Disc (HD-DVD)*
Capacidade de armazenamento: 15 GB (camada única)
30 GB (camada dupla)
- *Blu-Ray*
Capacidade de armazenamento: 27 GB (camada única)
54 GB (camada dupla)

Magnéticas

- *Hard Disk (HD)*
Capacidade de armazenamento: 2 TB
- Fita Magnética

As mídias ópticas geralmente são usadas para armazenamento de informações que não sofrerão alterações, as magnéticas destinam-se ao armazenamento de informações que necessitam de alterações em seu conteúdo.

1.6.4 Distribuição

A distribuição tem por finalidade garantir o acesso a determinados documentos. A função de distribuição também controla o acesso, ou seja, determina quais os documentos podem ser acessados dentro e fora da organização e quem terá acesso aos mesmos.

1.6.4.1 Tecnologias Relacionadas à Distribuição

a) *Enterprise Report Management (ERM)*

A tecnologia de *Enterprise Report Management (ERM)* tem a finalidade de gerenciar relatórios gerados em sistemas integrados da organização, geralmente com muitas páginas, como por exemplo, sistemas de *Enterprise Resource Planning*

(ERP) e *Customer Relationship Management* (CRM). Os sistemas de ERM permitem capturar arquivos diretamente na fila de impressão através de dispositivos que geram a imagem do documento e armazená-los em máquinas servidoras. Esses arquivos através de processos de indexação automática permitirão a localização de documentos por meio de palavras chaves de busca ou texto completo. Os sistemas de ERM permitem a consulta dos dados através de rede local ou internet.

Com a facilidade da internet, os clientes passaram a acessar os relatórios através da mesma. As principais justificativas para sua utilização são: redução no custo de impressão, disponibilização digital dos relatórios, agilidade no processo de consulta, etc. Aplicações típicas: faturas de telefone, energia elétrica, água, extratos bancários, relatórios financeiros entre outros.

1.6.5 Preservação

Os documentos digitais devem se manter acessíveis e utilizáveis pelo tempo que se fizer necessário. Deverá ser garantida sua longevidade, funcionalidade e acesso contínuo. Os principais agentes identificados como causadores de danos são: temperatura, umidade relativa do ar, tempo de uso da mídia, qualidade da mídia, manipulação das mídias, falhas no processo de fabricação e campo magnético.

Por outro lado, como o ambiente tecnológico está sempre em mudanças, sugere-se a adoção de diferentes estratégias para a preservação dos documentos digitais. Entre elas, destacam-se:

- **Arquivamento de Longo Prazo** – O microfilme ainda é um meio seguro para a preservação da informação. Se armazenado em condições adequadas, pode preservar seu conteúdo até 500 anos, conforme pode ser observado na Figura 1 (Cenadem, 2008e).

Nome da Mídia	Temp. °C	Umidade Relativa %	Durabilidade - Anos
CD-ROM	40	80	2
	30	60	10
	20	40	50
	10	25	200
WORM	40	80	5
	30	60	20
	20	40	100
	10	25	200
CD-R	40	80	2
	30	60	5
	20	40	30
	10	25	100
MAGNETO- ÓPTICO	40	80	2
	30	60	5
	20	40	30
	10	25	100
Microfilme com Qualidade Arquivística (Prata)	40	80	20
	30	60	50
	20	40	200
	10	25	500

Figura 1- Durabilidade das mídias
Mídias x Condições Ambientais x Durabilidade

Assim sendo, a associação da microfilmagem com o uso da tecnologia digital (soluções híbridas) tem se mostrado eficiente para resolver os problemas de preservação e acesso aos acervos documentais (WILLIS, 2001). Contudo o filme só pode ser lido na forma seqüencial. A utilização das duas tecnologias permite ao mesmo tempo garantir a preservação da imagem e o acesso. É aconselhável agregar o número do fotograma ao Banco de Dados em seu registro correspondente.

- **Migração** – A partir do envelhecimento da mídia, o conteúdo, deverá ser transferido para nova mídia para que possa continuar a ser acessado.
- **Back up** – O conteúdo deverá ser salvo em vários formatos e/ou locais, garantindo a viabilidade do negócio em caso de desastre. Deve-se ter garantia que as informações foram corretamente transferidas.
- **Utilização de formatos digitais abertos** – A adoção de formatos digitais abertos¹ também é uma medida de preservação recomendável e necessária.

¹Segundo o Serviço Federal de Processamento de Dados (Serpro), formato aberto trata-se de definir um padrão universal que pode ser aberto, editado e salvo em qualquer programa que se proponha a adotá-lo, como, por exemplo, editores de texto. Disponível em: <http://www.serpro.gov.br/noticias-antigas/noticias-2006/20060731_02>.

Conforme apresentado, o GED possui um grande leque de tecnologias. Segundo, Maurício Antonio Ferreira, especialista em GED e diretor de tecnologia do Cenadem, a utilização do GED e tecnologias correlatas em nosso País ainda causa muitas dúvidas, apesar de já termos instalações complexas e de alto nível.

Tratar as informações e documentos das empresas e organismos baseando-se em todas as tecnologias descritas que estão envolvidas pelo ECM ainda não é um procedimento concreto no Brasil, onde as companhias ainda estão descobrindo o *Imaging* como solução para eliminar papéis (Cenadem, 2004).

2 MUSEU PAULISTA DA USP

2.1 A INSTITUIÇÃO

O Museu Paulista da Universidade de São Paulo é mais conhecido como Museu do Ipiranga, pois se localiza próximo ao Riacho do Ipiranga, que viria a dar nome, também, ao bairro em seu entorno, na capital paulista. Faz parte do conjunto denominado Parque da Independência, especialmente formado como memorial nacional, por ser o local onde D. Pedro, em 7 de setembro de 1822, teria dado o “Grito”, símbolo da separação do Brasil de Portugal. Nele se encontram os jardins, a Casa do Grito, Monumento da Independência e a Capela Imperial Leopoldina. O Museu, inaugurado em 1895, é o mais antigo da cidade de São Paulo, incorporado à Universidade de São Paulo (USP), em 1963. É uma instituição científica, cultural e educacional, especializada no estudo dos aspectos materiais da sociedade brasileira. Suas exposições contam a história da sociedade entre os séculos XIX e XX. Entre suas peças destacam-se pinturas, esculturas, fotografias, móveis, louças, armas, automóveis distribuídos em seu interior. O acervo do Museu encontra-se tombado pelo Instituto do Patrimônio Histórico e Artístico Nacional (Iphan).



Figura 2 - Museu Paulista da Universidade de São Paulo

O Museu Republicano "Convenção de Itu" (MRCI) foi inaugurado em 1923, como instituto complementar do Museu Paulista. Em 1963, também foi incorporado à Universidade de São Paulo. Localizado no centro da cidade paulista de Itu, esse museu de caráter histórico dedica-se ao período conhecido como "República Velha", compreendido entre 1889 e 1930. O Museu Republicano possui um rico acervo de fotos, objetos e documentos entre outros. Devido a seu amplo acervo, o Museu, hoje, é uma referência de estudos, em todo o Brasil. Em 2005, foi cedida, pela Prefeitura de Itu, a Casa do Barão - casarão localizado a poucos metros do Museu Republicano, que tem como finalidade ser um local de estudo, pesquisa e realização de atividades educativas relacionadas a esse período (SAMARA, 2008).



Figura 3 - Museu Republicano "Convenção de Itu"



Figura 4 - Centro de Estudos do Museu, inaugurado em 2005

2.2 MISSÕES ESSENCIAIS DE UM MUSEU UNIVERSITÁRIO

O Museu Paulista define-se como uma instituição que produz conhecimento histórico na área de Cultura Material da sociedade brasileira, em especial paulista e, preferencialmente, no período entre 1850 e 1950. O MP, como instituição universitária, destina-se não apenas a coletar e guardar o acervo, mas sim compreender aspectos da sociedade que os produziu e utilizou. Seu acervo constitui-se na matéria prima necessária para a produção desse conhecimento. Entre suas atividades destacam-se:

a) Estudo do seu acervo

A partir de seu grupo de pesquisadores e especialistas, o Museu Paulista tem como responsabilidade o estudo de seus acervos, visando à geração de conhecimento.

b) Socialização do conhecimento

O Museu Paulista preocupa-se com a produção e difusão do conhecimento, por meio de catálogos (impressos ou eletrônicos), ações educativas e, especialmente, da promoção de exposições, seminários, cursos, palestras, visitas técnicas, estágios, etc. O MP também publica o periódico, *Anais do Museu Paulista*.

c) Preservação do acervo

O Museu Paulista está envolvido com a preservação física de seu acervo. Para tanto possui laboratórios especializados na conservação e restauração de papel, fotografia, tela e tecidos. Também no campo da preservação, o MP tem a função de produzir conhecimento.

d) Formação e ampliação de suas coleções

Por meio de incentivo a doações, aquisições ou coleta de campo de novas coleções, o Museu Paulista se preocupa em ampliar seu acervo em suas três linhas de pesquisa (Cotidiano e Sociedade, História do Imaginário e Universo do Trabalho).

Essas aquisições permitem a obtenção da matéria-prima necessária para produção do conhecimento.

2.3 ACERVO DO MUSEU PAULISTA DA USP

O acervo do Museu Paulista conta com objetos, material iconográfico e documentação arquivística. O acervo de objetos reúne mais de 26.000 unidades descritas em 26 categorias distintas. Entre elas: numismática, medalhística, filatelia, bandeiras, mobiliário e acessórios, indumentária, armaria, brinquedos, veículos, esculturas, entre outras.

O acervo iconográfico conta com mais de 50.000 itens. Inclui pinturas, desenhos, fotografias, gravuras, impressos, negativos, cartões postais, mapas e plantas. Entre suas coleções, destacam-se a de Militão Augusto de Azevedo, Santos Dumont, Bernardelli, Orôncio Vaz de Arruda, Aguirra e cartões postais referentes a cidades brasileiras, com maior ênfase da cidade de São Paulo. Podem também ser destacadas algumas pinturas sobre tela como: Pedro Américo (Independência ou Morte), Benedito Calixto (Inundação da Várzea do Carmo, 1892), Almeida Jr. (A Partida da Monção), entre outras.

O acervo textual reúne coleções e fundos de arquivos públicos e privados, além de documentos gerados na própria instituição entre 1893 e 1963, data da incorporação do Museu à Universidade de São Paulo. Hoje, totaliza aproximadamente 900 (novecentos) metros lineares de documentos, aproximadamente 300.000 documentos, reunidos em cerca de 110 (cento e dez) fundos e coleções. Entre eles destacam-se ofícios e correspondências, cartas pessoais, além de convites, menus, certidões, declarações, diplomas, contratos, procurações, inventários, escrituras, testamentos, etc. Conta também com uma hemeroteca que abriga, entre outros títulos, a coleção completa do jornal O Estado de São Paulo, desde 1875, e do jornal da tarde, desde 1966 (Witter, 1999).

2.4 ACERVO DO MUSEU REPUBLICANO CONVENÇÃO DE ITU

O acervo do MRCI é formado por objetos, documentos iconográficos e documentos textuais associados ao período conhecido como “República Velha”.

Seu acervo textual possui cerca de 400 metros lineares de documentação, aproximadamente 100.0000 documentos, abrangendo os períodos entre 1760 e 1932. É composto por 6 fundos e 33 coleções de origem pública ou privada, além de documentos gerados pela própria instituição (Zequini, 2000).

2.5 SISTEMA DOCUMENTAL INFORMATIZADO NO MUSEU PAULISTA

Tendo em vista uma atualização das práticas arquivísticas, assim como uma maior eficiência no acesso de pesquisadores a seus acervos, o corpo técnico-científico do Museu Paulista, em 1991, iniciou projeto que propunha soluções associadas à tecnologia da informação. Optou-se, naquela época, pela utilização de uma ferramenta de desenvolvimento, de largo uso comercial e baixo custo de aquisição e licenciamento, tendo em vista a formação de um Banco de Dados.

Foi desenvolvido pelo seu Analista de Sistemas, um programa para atender ao acervo iconográfico e de objetos da instituição. Inicialmente utilizando-se a linguagem de programação *FoxPro*, passando em seguida, para o *Visual FoxPro* 6.0 até chegar à versão 9.0, atualmente utilizada. O programa possibilita a inserção de um número ilimitado de imagens para cada item catalogado. Também permite a localização do acervo através do cruzamento de vários campos, além de possibilitar a emissão de relatórios diversos. Contudo, não possui interface WEB, o que não permite disponibilizar seu acervo na Internet. Atualmente esse sistema possui 39.521 itens do acervo iconográfico e 22.106 itens de seu acervo de objetos catalogados.

MUSEU PAULISTA DA UNIVERSIDADE DE SÃO PAULO

Cadastro Acessórios Relatórios Administração

Resumo 1 2 3 4 Inventário Descritores Galeria

Número (IC) 1-22823-0000-0000

Denominação FOTOGRAFIA/FOTOPINTURA

Título/Legenda RETRATO EM TAMANHO NATURAL DE SANTOS DUMONT*

Autoria1 SARRACINO, GIOVANNI

Autoria2 DESCONHECIDA

Século 20 **Década** 1900 **Data** / /1904

Datas-limite - **Observações** ATRIBUÍDO

Origem SÃO PAULO/BRASIL

Ref./Agente SARRACINO, RUA FORMOSA,85 (1901- ?)

Mat./Técnica PAPEL, TECIDO, GELATINA, NITRATO PRATA, AQUARELA

Fundo/Coleção ALBERTO SANTOS DUMONT - CSD

N. MP 25276 **N. RUSP** 180359 **N. Mercúrio** 0

Acesso

LIVRE

RESTRITO

NÃO RECOMENDADO

CONSULTE

Início Anterior Próximo Último

Novo Edita Reverte Grava Apaga Localiza Filtra Grade Imprime Sai

Ordem

Nenhuma

1-22823-0000-0000



Figura 5 – Tela do sistema (módulo de iconografia)

A presente proposta está voltada para o acervo textual do Museu Paulista, o qual não conta ainda com nenhum sistema informatizado. Pretende-se, com a informatização desse acervo, integrá-lo aos demais acervos já informatizados, criando-se um sistema único de gerenciamento de documentos que estará disponível para consulta por meio da rede mundial de computadores.

Assim, vem ao encontro dos esforços da Direção e dos docentes-pesquisadores, preocupados em responder com eficiência e rapidez às demandas, sempre crescentes, da pesquisa universitária, assim como atender aos interesses de diferentes públicos e instituições.

3 IMPLANTAÇÃO DE PROJETO DE GED NO MUSEU PAULISTA: METODOLOGIA, DIRETRIZES E EXPERIMENTAÇÃO.

3.1 METODOLOGIA APLICADA

Para implantar o GED, é importante a utilização de metodologia adequada, de forma a se ter controle sobre as variáveis envolvidas.

Um referencial no mercado é a proposta desenvolvida pelo *Project Management Institute* (PMI)² que descreve um conjunto de conhecimentos e práticas aplicáveis à maioria dos projetos. Todas as informações relacionadas a Gerenciamento de Projetos estão definidas no manual do PMI, “Um guia de Conhecimentos em Gerenciamento de Projetos”, conhecido também como Guia PMBOK.

Já o e-ARQ Brasil estabelece oito passos que podem ser desenvolvidos em diferentes estágios, interativamente, parcial ou gradualmente, de acordo com as necessidades do órgão ou entidade. A metodologia prevê, ainda, ciclos de aplicação, sendo que as tarefas previstas do passo 1 ao 8 devem ser realizadas periodicamente (CTDE, 2006).

Os oito passos referidos são:

- 1) Levantamento Preliminar;
- 2) Análise das funções, das atividades desenvolvidas e dos documentos produzidos;
- 3) Identificação das exigências a serem cumpridas para a produção de documentos;
- 4) Avaliação dos sistemas existentes;
- 5) Identificação das estratégias para satisfazer as exigências a serem cumpridas para a produção de documentos arquivísticos;
- 6) Projeto do sistema de gestão arquivística de documentos;
- 7) Implementação do sistema de gestão arquivística de documentos;
- 8) Monitoramento e ajustes.

² Ver em: <<http://www.pmi.org>>. Acesso em 12 jan. 2009.

Uma terceira metodologia desenvolvida pelo *The Rheinner Group*³ e apresentada por Haddad em sua monografia, detalha as etapas e fases de tratamento de documentos (HADDAD, 2000).

Apresenta-se dividida em cinco fases:

- 1) Descoberta
- 2) Levantamento
- 3) Análise
- 4) Projeto
- 5) Construção

Segundo *The Rheinner Group* (apud HADDAD,2000), essa metodologia apresenta um ciclo entre suas fases, ou seja em cada ciclo poderá haver um reinício na primeira fase.

O fato é que não existe uma metodologia padronizada para aplicação do GED. Também não se pode afirmar que uma metodologia é melhor ou pior que a outra. Embora as metodologias sejam apresentadas em formato diferente, segundo Baldam (2002), todas as metodologias possuem os itens que seguem:

- Definição da Solução
- Desenvolvimento da Solução
- Funcionamento da Solução
- Avaliação e Melhoria da Solução

Nesse trabalho será usada a metodologia apresentada pelo *The Rheinner Group*, agrupando-se os itens Levantamento e Análise, por estarem fortemente relacionados. Essa metodologia é recomendada pela sua simplicidade e estar praticamente moldada às necessidades identificadas no Museu Paulista.

O objetivo deste estudo não é apenas descrevê-la, mas iniciar o desenvolvimento de um projeto piloto. Serão levantados problemas e propostas possíveis soluções. Entretanto, o projeto não poderá ser concluído, pois depende de resoluções a serem tomadas em conjunto com os membros da equipe, tais como: prazo para a realização do projeto, recursos financeiros disponíveis, campos que

³ *The Rheinner Group* é uma entidade de pesquisa, ensino e consultoria em DI, DM e *Workflow*. São responsáveis pelo programa de treinamento e certificação nas tecnologias de GED, denominado *Certified Document Imaging Architect* (CDIA).

farão parte do sistema informatizado, campos a serem indexados, acervo participante do projeto, quantidade de documentos existentes e seus formatos, quem terá acesso a esses documentos, requisitos de software, requisitos para a digitalização das imagens, requisitos de desempenho, entre outros. A partir dessas informações, poderão ser levantados os custos, recursos humanos necessários, equipamentos e definido um cronograma do projeto.

3.1.1 **Descoberta**

A partir das primeiras informações obtidas acerca do GED apresentadas no primeiro capítulo, será necessário realizar novas pesquisas de aprofundamento. Isso poderá se dar através da participação dos profissionais envolvidos em cursos, palestras, realização de eventos, congressos, seminários etc. Também poderão ser efetuadas visitas a empresas correlatas que implantaram o GED, visando tomar conhecimento do processo e principais dificuldades na implantação.

Entre outros aspectos, esses estudos se fazem necessários, pois o próprio *The Rheinner Group* (apud HADDAD, 2000) publicou que um dos principais motivos para o insucesso de projetos envolvendo tecnologias do GED relaciona-se com a estratégia das organizações de adquirirem as soluções antes de conhecerem suas necessidades.

3.1.2 **Levantamento / Análise**

Após a fase da descoberta teve início o levantamento dos problemas encontrados no departamento a ser informatizado, no caso o SVDHICO. A partir do conhecimento dessas dificuldades e dos benefícios esperados, poderão ser definidas as tecnologias do GED aplicáveis ao projeto.

O passo seguinte será a identificação dos recursos humanos e equipamentos disponíveis para a execução do projeto, com o objetivo de diagnosticar a situação atual. Durante o processo de levantamento e análise, também serão identificados

requisitos necessários à implantação do GED. Somente a partir da resolução dessas questões é que ocorrerá o efetivo detalhamento do projeto.

a) Problemas diagnosticados no Serviço de Documentação

Dentre as questões já levantadas, percebeu-se que, embora os documentos estejam devidamente armazenados e ordenados através de um plano de classificação (guia das coleções, catálogo em forma de fichas descrevendo documento a documento), a localização desses documentos no acervo é feita de forma manual.

Por outro lado, apesar de todas as medidas tomadas para a preservação dessa documentação, é fato que a manipulação desses documentos contribui com sua deteriorização, além de sujeitá-los a possíveis danos ou furtos.

Outro aspecto considerado é que a busca é feita documento a documento, não havendo a possibilidade de se acessar as informações de forma rápida e eficiente. Desta maneira torna-se difícil democratizar o acesso, bem como preservar os documentos históricos. Entretanto, em função da própria natureza do equipamento cultural, não se pode impedir o acesso dos consulentes a essa massa documental.

A seguir, em função dos principais problemas enfrentados pelo Museu Paulista da USP em relação à sua documentação textual serão arrolados os problemas e demonstradas as implicações e benefícios a serem alcançados com a implantação do Gerenciamento Eletrônico de Documentos (GED).

Descrição do Problema	Implicação	Benefícios obtidos com a implantação do GED
Manipulação de Documentos.	Aceleração do processo de deterioração dos documentos.	Diminuição da manipulação e conseqüente aumento na preservação dos documentos
Localização manual dos documentos.	Dificuldade na localização de forma rápida e eficiente.	Ilimitadas possibilidades para indexação e alta velocidade e precisão na localização de documentos.

continua

continuação

Descrição do Problema	Implicação	Benefícios obtidos com a implantação do GED (Tecnologia DI)
Restrição quanto à tiragem de cópias.	Dificuldade de comprovação das informações.	Possibilidade de impressão de documentos sem restrição e sem prejuízo à conservação.
Falta de espaço Físico e Recursos Humanos para a disponibilização de documentos.	Dificuldade na disponibilização dos documentos. Impede a democratização do acervo.	Disponibilização instantânea de documentos sem limites físicos e recursos humanos. Melhor atendimento ao consulente.
Disponibilização do acervo somente na instituição.	Dificuldade de acesso à informação por consulentes de outras localidades.	Possibilita o acesso remoto baseado em navegadores.
Possibilidade de furto de documentos ou desastre.	Perda dos documentos	Produção de cópia do documento. Documentos consultados sem riscos de serem furtados ou danificados. Preservação da informação em caso de desastre.

Figura 6 – Problemas, Implicações e benefícios com a implantação do GED

A partir do entendimento das tecnologias existentes e da identificação dos problemas da instituição, percebe-se que o *Document Imaging* é a mais adequada, porém outras tecnologias complementares poderão ser utilizadas, por exemplo, o OCR e ICR, conforme o resultado obtido.

b) Diagnóstico dos Recursos Humanos existentes

A implantação do GED pressupõe levantamento dos recursos humanos disponíveis e seus níveis de habilidade para verificar se esses recursos atendem ao projeto.

Serviço de Documentação Textual e Iconografia:

O serviço de documentação possui conhecimento arquivístico de toda a documentação textual e iconográfica. É o responsável pela guarda e disponibilização deste acervo. As consultas, no momento, só podem ser efetuadas através de agendamento prévio. Atualmente são atendidos 500 consulentes ao ano, porém muitos acabam desistindo pela demora no agendamento. Essa demora se dá principalmente devido à falta de funcionários no setor e limite do espaço físico.

O setor conta atualmente com dois especialistas em pesquisa e um técnico de museu. Percebe-se que só há disponibilidade efetiva de pessoal para a coordenação do projeto, visto que estes funcionários já desempenham atividades na instituição.

Seção Técnica de Informática:

A seção técnica de informática é atualmente composta por um analista e dois técnicos de informática. Responde pela implantação e manutenção de todo ambiente tecnológico da instituição, inclusive telefonia, sistemas informatizados para a segurança do acervo, catracas eletrônicas, controle de acesso, áudio guias entre outros. Mais uma vez, só há disponibilidade efetiva de pessoal para a coordenação do projeto, visto que estes funcionários já desempenham inúmeras atividades na instituição.

c) Diagnóstico dos Recursos Tecnológicos Existentes

Os Recursos Tecnológicos existentes, tais como computadores, servidores, scanners, impressoras, ambiente operacional e a infra-estrutura, deverão ser levantados a fim de se verificar se são suficientes e atendem ao projeto.

Equipamentos:

O MP e MRCl contam com aproximadamente 130 microcomputadores além

de periféricos distribuídos em suas dependências, contudo serão considerados apenas os equipamentos disponíveis para a implantação do projeto, ou seja: 10 microcomputadores, 1 impressora, 1 servidor e 2 scanner de mesa. O ambiente operacional em uso é o *Windows*.

Infra-estrutura de Rede:

O Museu Paulista dispõe, para acesso externo, de uma rede *Frame Relay*⁴ de 2 megabits por segundo (Mbps), compartilhada com o Museu de Zoologia da Universidade de São Paulo, que fica localizado nas imediações do Museu Paulista. A interligação entre esses dois edifícios é feita através de fibra óptica. Os microcomputadores da instituição estão interconectados através de rede cabeada (100 Mbps) e rede sem fio (*Wireless*) 802.11G (54 Mbps) e 802.11N (300 Mbps).

Conhecer os equipamentos e a infra-estrutura existentes se faz necessário para avaliar as condições tecnológicas do setor a ser informatizado, a fim de verificar se estas condições são adequadas à necessidade do projeto.

d) Questões a serem resolvidas

Critérios para a escolha do formato de gravação:

Deverão ser avaliados os formatos de gravação existentes para que, quando for efetuada a digitalização das imagens, já se tenha decidido qual o formato de gravação que será utilizado. Os formatos mais adequados são os amplamente aceitos como o *Tagged Image File Format* (Tiff), *Joint Photographic Experts Group* (Jpeg) e *Portable Document Format* (PDF).

Devem ser levados em consideração os seguintes critérios para a escolha do formato de gravação:

- preferir formatos com especificações abertas ao domínio público;
- utilizar formatos de gravação padronizados e amplamente utilizados pela maioria dos fabricantes de dispositivos e *softwares*.

⁴ O *Frame Relay* é uma tecnologia de comunicação de dados de alta velocidade, usada para enviar informação digital através de uma rede de dados, dividindo essas informações em *frames* (quadros) ou *packets* (pacotes).

- preferir formatos estáveis (não sujeitos a constantes atualizações e compatíveis com versões anteriores).

Formato Tiff

O formato Tiff ou TIF é um formato de armazenamento que gera arquivos grandes e de alta qualidade. Pode também utilizar o algoritmo de compressão Lempel-Ziv-Welch (LZW) para reduzir o tamanho do arquivo em *bytes*. A compressão LZW não é destrutiva, ou seja, a imagem não perde sua definição original, contudo não tem um alto poder de compressão como o Jpeg. O formato Tiff, pertence hoje a Adobe Systems Inc; suas especificações estão disponíveis no site da Adobe <<http://www.adobe.com>>.

Formato Jpeg

O formato Jpeg é um formato aberto, padrão na internet e, apesar de usar a codificação destrutiva para a compressão das informações, prevê considerável economia de espaço em disco e facilidade de transmissão. A codificação Jpeg trabalha com níveis distintos de compressão destrutiva; quanto maior a compressão, maior será a destruição da informação e menor será o tamanho da imagem.

O algoritmo de compressão do Jpeg causa perdas que não podem ser mais recuperadas. A taxa de compressão não é fixa, áreas contínuas de mesma cor terão uma compactação maior. Portanto imagens de um mesmo tamanho em papel, após digitalizadas, poderão resultar em tamanhos de imagem diferentes. As especificações desse formato estão disponíveis no endereço eletrônico <www.jpeg.org>.

Formato PDF

Criado pela Adobe Systems, o formato *Portable Document Format* (PDF) é um padrão internacional de distribuição de documentos. Ele permite capturar e visualizar informações a partir de quase todo aplicativo ou sistema operacional e compartilhá-las praticamente com qualquer pessoa, em qualquer lugar. Através dele é possível assinar digitalmente ou proteger por senha documentos PDF. A partir de

2008, tornou-se um padrão aberto formal, conhecido como ISO 32000-1:2008. As especificações desse formato estão disponíveis no site da Adobe <<http://www.adobe.com>>.

Recomendações acerca dos Formatos

O Museu Paulista identifica a necessidade de se criar um Banco de Dados em alta definição com o propósito de preservar seus documentos em forma digital. Para esse propósito, o formato Tiff com compressão LZW parece ser o mais adequado. A compressão LZW não é destrutiva, contudo não tem um alto poder de compressão como o Jpeg, entretanto o formato Tiff não pode ser disponibilizado na Internet.

No entanto, como já se apresenta previsão da disponibilização da informação para pesquisas internas (rede local) e externas (Internet), o formato Jpeg parece ser o mais adequado para a disposição de arquivos de imagem. Ele permite um alto nível de compressão adequado à Internet e para a pesquisa como referência no Banco de Dados.

A grande vantagem do formato PDF é a possibilidade de criar arquivos, multipaginados, multiplataforma e pesquisáveis. No caso de pesquisa em textos (*Full Text*) só se aplica se for possível à conversão do arquivo digitalizado em formato de imagem para a forma de texto, através da utilização de tecnologias como o OCR/ICR.

A partir dessas análises verifica-se que há uma grande quantidade de formatos disponíveis no mercado. A adoção de um ou outro formato depende do resultado a ser obtido. No caso do Museu Paulista verifica-se que será necessário a adoção de diferentes formatos.

Cores:

A digitalização dos documentos em preto em branco ou em escala de cinza atenderia o requisito informação e traria benefícios como à redução substancial no tamanho da imagem gerada. Contudo, por tratar-se de documentos históricos, recomenda-se a digitalização em cores, ainda que a partir de uma seleção de documentos, para que seja mantida a maior fidelidade com os originais.

Resolução de digitalização:

A resolução determina o nível máximo de detalhamento que pode ser capturado em uma imagem. A resolução óptica de um scanner é medida pela capacidade de leitura do sensor óptico.

Segundo BASTOS⁵ (2009), a resolução de entrada ou de digitalização deverá ser a mesma resolução do dispositivo de saída, mantida as dimensões do documento original.

As impressoras fotográficas, como os modernos *minilabs* digitais, têm uma resolução média de 300 DPI (pontos por polegada). O mais sensato é optar pela impressão fotográfica, na qual o resultado é sempre uma fotografia de tom contínuo, não composto por retículas ou pontos de impressão visíveis.

Do ponto de vista da preservação e tratando-se de documentos textuais históricos, recomenda-se a digitalização em cores no formato Tiff, com resolução de 300 DPI para assegurar a reprodução das nuances, as condições do papel e outras marcas nele gravadas. As imagens digitalizadas em Tiff deverão ser convertidas para o formato Jpeg a fim de se disponibilizar na Internet.

Equipamentos a serem utilizados para a digitalização:

Tratando-se de livros ou documentos encadernados, recomenda-se a digitalização, por meio da utilização de scanner planetário ou câmeras digitais. Através da utilização desses equipamentos, não há a necessidade de desmontar os originais, preservando sua integridade. A digitalização utilizando o conjunto composto pelo sistema de iluminação, câmera e computador portátil; possui a vantagem do deslocamento com facilidade e obtenção de imagens de qualidade equivalente as do scanner planetário em condições controladas.

Tratando-se de documentos históricos não deverá ser utilizado scanner com dispositivo de tração automática de papel. No caso de documentos encadernados, o scanner de mesa não deverá ser adotado pela possibilidade de danos a lombada desses documentos.

⁵ Jorge Bastos, especialista em sistemas de imagens digitais é sócio-gerente da empresa Motivo Processamento, Imagem e Comunicação Ltda. Dirigiu, desde a década de 1990, empresas de tecnologia em reprodução de imagens no Brasil, na Europa e nos Estados Unidos, recebendo prêmios e reconhecimento internacional em sua área de atuação.

Espaço em disco necessário para armazenagem de documentos:

Para descobrir com exatidão o tamanho necessário para o armazenamento das imagens digitais, recomenda-se a digitalização de amostras de documentos de vários tamanhos a fim de determinar o tamanho aproximado de cada imagem.

A partir da amostragem, multiplicar o tamanho de cada documento pela quantidade de documentos, a fim de se estimar o tamanho de disco necessário para armazenagem da documentação. Segundo Baldam (2002), deve-se utilizar uma margem de segurança de 30% sobre o total de armazenagem.

Também é possível dimensionar o tamanho das imagens por meio da utilização de fórmulas. Mas tratando-se de arquivos comprimidos, o tamanho do arquivo pode variar de documento a documento. Essa variação se dá devido o algoritmo de compressão atuar de forma diferente nas diversas áreas do documento.

Qualidade das imagens digitais:

A avaliação da qualidade das imagens digitais geradas tem por finalidade garantir que os padrões selecionados sejam seguidos e os resultados, os previstos. Para avaliar a qualidade da imagem, devem ser selecionadas algumas imagens por amostragem, preferencialmente no início, meio e fim de cada lote, a fim de verificar se:

- a imagem exhibe o original inteiro;
- a imagem está torta;
- o texto é legível;
- a imagem está focada;
- a cor está correta ;
- a tarja de controle está presente na imagem;
- a resolução está dentro do esperado;
- o enquadramento está correto.

Recomenda-se a utilização de um editor de imagens que apresentem histogramas para avaliação da qualidade das imagens digitais. Um histograma é