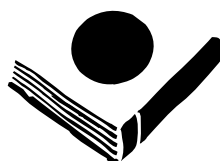


Do microfilme à imagem digital

Donald J. Waters

2ª edição



CONSERVAÇÃO PREVENTIVA
EM BIBLIOTECAS E ARQUIVOS

Donald J. Waters

**Do microfilme à imagem digital:
como executar um projeto para estudo dos
meios, custos e benefícios de conversão
para imagens digitais de grandes
quantidades de documentos
preservados em microfilme**

2ª edição

Rio de Janeiro
Projeto Conservação Preventiva em Bibliotecas e Arquivos
2001

Copyright ©1991 by The Commission on Preservation and Access.

Título original, publicado por *The Commission on Preservation and Access*:
From Microfilm to digital imagery: on the feasibility of a project to study the means, costs and benefits of converting large quantities of preserved library materials from microfilm to digital images.
Autor: Donald J. Waters

Projeto cooperativo interinstitucional Conservação Preventiva em Bibliotecas e Arquivos, em parceria com o CLIR - Council on Library and Information Resources (Conselho de Recursos em Biblioteconomia e Informação, que incorporou a antiga Commission on Preservation and Access).

Suporte Financeiro
The Andrew W. Mellon Foundation
Vitae, Apoio à Cultura, Educação e Promoção Social

Apoio
Arquivo Nacional
Fundação Getulio Vargas

Coordenação
Ingrid Beck

Colaboração
Sérgio Conde de Albite Silva

Tradução
José Luiz Pedersoli Júnior

Revisão Técnica
Mauro Resende de Castro
Ana Virginia Pinheiro
Dely Bezerra de Miranda Santos

Revisão Final
Cássia Maria Mello da Silva
Lena Brasil

Projeto Gráfico
T'AI Comunicações

Coordenação Editorial
Ednéa Pinheiro da Silva
Anamaria da Costa Cruz

Impresso em papel alcalino.

W329 Waters, Donald J.

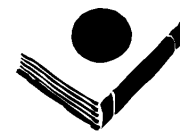
Do microfilme a imagem digital : como executar um projeto para estudo dos meios, custos e benefícios de conversão para imagens digitais de grandes quantidades de documentos preservados em microfilme / Donald J. Waters ; [tradução José Luiz Pedersoli Júnior ; revisão técnica Mauro Resende de Castro, Ana Virginia Pinheiro, Dely Bezerra de Miranda Santos; revisão final Cássia Maria Mello da Silva, Lena Brasil] . - 2. ed. - Rio de Janeiro : Projeto Conservação Preventiva em Bibliotecas e Arquivos: Arquivo Nacional, 2001.

43 p. : il. ; 30 cm. - (Conservação Preventiva em Bibliotecas e Arquivos; 49. Reformatação).

Inclui bibliografias.
ISBN 85-7009-053-6.

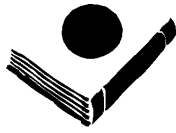
1. Microfilmagem. 2. Reprografia. I. Título. II. Série.

CDD 686.43



Sumário

Apresentação	5
Prefácio da <i>Commission on Preservation and Access</i>	7
Introdução	9
Visão	10
Preservando o acesso através da preservação	14
Uma economia de escolhas	18
Arquitetura do sistema	25
Plano de trabalho	34
Conclusão	42
Notas	43



Apresentação

O Projeto *Conservação Preventiva em Bibliotecas e Arquivos - CPBA* é uma experiência de cooperação entre instituições brasileiras e a organização norte-americana *Commission on Preservation and Access*, atualmente incorporada ao *CLIR - Council on Library and Information Resources* (Conselho de Recursos em Biblioteconomia e Informação).

Em 1997, o Projeto traduziu e publicou 52 textos sobre o planejamento e o gerenciamento de programas de conservação preventiva, onde se insere o controle das condições ambientais, a prevenção contra riscos e o salvamento de coleções em situações de emergência, a armazenagem e conservação de livros e documentos, de filmes, fotografias e meios magnéticos; e a reformatação envolvendo os recursos da reprodução eletrônica, da microfilmagem e da digitalização.

Reunidos em 23 cadernos temáticos, estes textos, somando quase mil páginas, foram impressos com uma tiragem de dois mil exemplares e doados a colaboradores, instituições de ensino e demais instituições cadastradas no banco de dados do Projeto.

Esta segunda edição revisada, com uma tiragem de mais dois mil exemplares, pretende, em continuidade, beneficiar, as instituições e os profissionais de ensino, e todas aquelas instituições inscritas no banco de dados depois de 1997 e que não chegaram a receber os textos.

O presente caderno, de número 49, relata aspectos de um projeto de conversão de microfilme em imagem digital, desenvolvido nos Estados Unidos. Aborda não apenas a tecnologia, mas a preocupação de assegurar a preservação da informação, o seu uso e acesso. Apresenta um plano de trabalho para desenvolver sistemas práticos para a conversão de microfilme em imagem digital. Mapeia a arquitetura de um sistema de reprodução digital de imagens, identificando pontos críticos. Ao final apresenta o escopo do projeto aplicado na Universidade de Yale, em uma biblioteca de 10 mil livros, digitalizados a partir de cópias em microfilme.

Este texto, assim como todo o conjunto de publicações do Projeto CPBA, encontra-se disponível em forma eletrônica na página do Projeto, www.cpba.net.

Além das publicações distribuídas em 1997, o Projeto CPBA ainda formou multiplicadores, por meio de seminários organizados nas cinco regiões brasileiras, com o apoio de instituições cooperativas. Os multiplicadores organizaram novos eventos, estimulando a prática da conservação preventiva nas instituições. No início de 2001 o projeto já contabilizava mais de 120 eventos realizados, somando mais de quatro mil pessoas envolvidas. Os inúmeros desdobramentos ocorridos a partir dos colaboradores em todo o país fizeram o Projeto merecedor, em 1998, do Prêmio Rodrigo Melo Franco de Andrade.

Entre 1997 e 2001, o Projeto CPBA continuou promovendo seminários e cursos, envolvendo as instituições cooperativas. Em muitas ocasiões enviou professores e especialistas aos eventos organizados pelos multiplicadores. No início de 2001 o Projeto já contabilizava mais de 120 eventos realizados em todo o país, somando mais de 4.000 pessoas envolvidas.

As instituições que colaboram com o Projeto CPBA estão relacionadas na página www.cpba.net, onde também poderá ser acessado o seu banco de dados, com mais de 2.600 instituições cadastradas. Esta página virtual pretende ser uma plataforma para o intercâmbio técnico e o desenvolvimento de ações cooperativas.

Desde o início o Projeto contou com recursos financeiros da *Andrew W. Mellon Foundation* e de VITAE, Apoio à Cultura, Educação e Promoção Social. Em 1998 estes patrocinadores aprovaram um segundo aporte financeiro, com o objetivo de dar continuidade às ações empreendidas e de preparar esta segunda edição.

O Projeto agradece o generoso apoio recebido de seus patrocinadores e das instituições cooperativas, brasileiras e estrangeiras, reconhecendo que sem esta parceria nada teria acontecido. Deseja também agradecer aos autores e editores das publicações disponibilizadas, por terem cedido gratuitamente os direitos autorais. Agradecimentos especiais ao Arquivo Nacional, que hospedou o Projeto desde o seu início, assim como à Fundação Getúlio Vargas, pela administração financeira dos recursos.

Considerando que a fase do Projeto apoiada pela Fundação Mellon se encerra em junho de 2001, o grupo cooperativo espera encontrar, em continuidade, colaboradores e parceiros no Brasil, para que o processo de difusão do conhecimento da preservação não seja interrompido.

Rio de Janeiro, junho de 2001.

Ingrid Beck
Coordenadora do Projeto CPBA

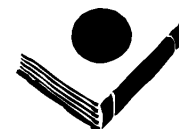


Prefácio da *Commission on Preservation and Access*

Em um relatório da *Commission*, Michael Lesk, membro do *Technology Assessment Advisory Committee*, declara que é recomendável, a longo prazo, a conversão de microfilme para imagem digital objetivando a preservação e o acesso de documentos em deterioração, sendo tecnicamente possível e relativamente barata de se realizar (*Image formats for preservations & access*, julho de 1990, página 8). Contudo, a tecnologia de imagem digital está ainda em desenvolvimento, os padrões para *hardware* e *software* continuam a evoluir e a conversão de microfilme para formatos de imagem digital ainda não está amplamente testada nas bibliotecas.

No estágio atual, é essencial direcionar a pesquisa e o investimento ao desenvolvimento de sistemas para gerenciamento da conversão digital, armazenamento e acesso, para que possam ser utilizados por universidades e bibliotecas de pesquisa. É para o desenvolvimento de tais sistemas que a Biblioteca da Universidade de Yale direcionou seus esforços neste estudo de planejamento.

O estudo, encomendado pela *Commission*, explora a exequibilidade de um projeto para estudar os meios, custos e benefícios da conversão para imagens digitais de grandes quantidades de documentos de bibliotecas, preservados em microfilme. O Estudo identifica as necessidades visando um projeto maior, de anos de duração, para proporcionar acesso tanto intra quanto inter-institucional às imagens armazenadas e para investigar as implicações mais gerais da intensificação do acesso intelectual à documentação digitalizada. A *Commission* espera que os resultados se constituam numa importante contribuição à nossa capacidade de utilização das novas tecnologias digitais para propósitos de preservação. Cópias gratuitas das publicações deste relatório foram distribuídas aos assinantes da *Commission*.



Introdução

Nas últimas três décadas, o processo de microfilmagem foi considerado o mais eficiente para a preservação de documentos em deterioração nas bibliotecas: o microfilme é durável, desde que armazenado num ambiente controlado e que bibliotecas e arquivos sigam procedimentos e especificações normatizados, podendo partilhar até mesmo instalações comuns na microfilmagem. Os custos do processo são previsíveis, variando de 10 a 15 centavos de dólar por página, excluindo-se os custos de despesas gerais administrativas e da seleção do material a ser microfilmado. Além disso, a tecnologia para se ter acesso ao microfilme é estável e dificilmente sofrerá qualquer alteração significativa no futuro. Em resumo, para a solução do problema dos livros quebradiços, que atualmente afeta todos os depósitos de material impresso sobre papel ácido, o microfilme é essencialmente livre de riscos¹.

Michael Lesk, gerente da Divisão de Pesquisa em Ciência da Computação na Bellcore, em um recente relatório apresentado à *Commission on Preservation and Access*, ratificou a utilização do microfilme para preservação nos dias atuais. Contudo, ele foi relativamente cauteloso no tocante à sua visão sobre o papel do microfilme no futuro, argumentando que o mesmo, provavelmente, não é a última etapa no processo de preservação. Mais propriamente, argumentou ele, o microfilme é “um razoável passo intermediário para a obtenção de imagens digitais”, que é o processo de escaneamento de livros para armazenagem em computadores².

Apesar de Lesk ter admitido que a tecnologia de reprodução digital é cara e que “ainda não está bem estabelecida”, ele argumentou fortemente que esta tecnologia encerra uma promessa fundamental à missão das bibliotecas de gerar, preservar e otimizar o acesso ao conhecimento registrado. Ele observou que a conversão de microfilme para imagem digital é tecnicamente possível hoje e argumentou que os usuários de bibliotecas irão, a longo prazo, exigir os resultados de tal conversão. Previu, ainda, que a tecnologia de imagem digital, incluindo a possibilidade de conversão a partir do microfilme, “estará ao alcance da maioria das bibliotecas ainda nesta década” e incentivou o desenvolvimento de pesquisas e investimentos, hoje, para o “desenvolvimento de sistemas que possam ser utilizados por bibliotecas comuns” no futuro³.

A biblioteca da Universidade de Yale, uma das primeiras instituições a desenvolver práticas de microfilmagem para preservação, está atualmente envolvida nos estágios iniciais de um esforço amplo e de longo prazo visando a desenvolver sistemas práticos para a conversão de microfilme em imagem digital e, assim, compreender melhor a relação adequada entre as duas tecnologias na prática de preservação futura. Este relatório resume os resultados da fase inicial de planejamento do projeto maior. Na próxima seção, começa-se a articular uma visão do lugar a ser ocupado pela tecnologia digital em uma biblioteca de pesquisa como a de Yale. Em seguida, faz-se um esboço de economia de

¹ Nancy E. Gwinn (Ed.). *Preservation microfilming: a guide for librarians and archivists*. Chicago: American Library Association, 1987. p. xxxvi; John C. Mallinson. On the preservation of human and machine-readable records. *Information Technology and Libraries*, v. 7, no. 1, p. 22, Mar. 1988.

² Michael Lesk. Digital imagery, preservation and access. *Information Technology and Libraries*, v. 9, no. 4, p. 307, Dec. 1990.

³ Idem, *ibidem*, p. 308.

escolhas e organização na qual a conversão de microfilme para imagem digital pode surgir como uma opção viável de preservação. O relatório mapeia a arquitetura de um sistema de reprodução digital de imagens e identifica pontos críticos de exequibilidade. Finalmente, o relatório apresenta um plano de trabalho para um projeto de demonstração, que investigará em profundidade os custos e benefícios da conversão no processo de construção de uma biblioteca digitalizada de 10 mil livros a partir de cópias em microfilme.

Visão

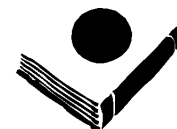
Para uma visão ampla do futuro papel das bibliotecas de pesquisa na universidade, deve-se olhar cuidadosamente, e de perto, a maneira como os estudiosos trabalham para produzir e disseminar o conhecimento. Uma série de estudos recentes enfatiza que a qualidade e a quantidade da produção dos estudiosos em todas as disciplinas dependem decisivamente do acesso pronto e oportuno à informação relevante. A biblioteca do futuro, para estes pesquisadores, não será necessariamente uma ‘biblioteca eletrônica’, nem será sua a missão de preservar simplesmente o conhecimento registrado. Ela terá, sim, como objetivo, mais apropriadamente, a geração, a preservação e a melhoria para seus usuários do pronto acesso — físico e intelectual — ao conhecimento registrado. Tecnologias emergentes, como a de imagem digital, terão então uma posição fundamental, na medida em que ajudam a sustentar a missão geral da biblioteca.

O Valor do Acesso

Iniciando em 1985, o *Research Libraries Group* estabeleceu um Programa Especial para Gerenciamento da Informação de Pesquisa, PRIMA. O programa resultou em uma avaliação detalhada das necessidades de informação em mais de 20 disciplinas de três grandes áreas de conhecimento: Ciências Humanas, Ciências Sociais e Ciências Físicas. Entre as várias necessidades que emergiram dessas avaliações, muitas eram específicas das disciplinas em estudo. Um tema, contudo, percorre todos os estudos e aplica-se a todas as disciplinas: os estudiosos, usando as palavras expressas no relatório sobre as necessidades de informação na área de humanas — “preferem os métodos familiares e não sistemáticos de consulta às notas de pé de página e aos colegas”⁴. Nenhuma evidência é fornecida nos relatórios para justificar a caracterização destes métodos como “não sistemáticos”, mas está claro que, para os pesquisadores, o valor de se consultar as notas de pé de página e os colegas está no fato de que esses métodos são prontos e facilmente acessíveis.

Um estudo patrocinado pelo *Faxon Institute for Advanced Studies in Scholarly and Scientific Communication*, apresentado numa conferência em abril de 1991, aborda o mesmo ponto, com respeito aos cientistas, de forma ainda mais intensa. O estudo examinou o comportamento de um número significativo de profissionais da área científica com relação à forma como eles adquiriam e utilizavam informações em seu trabalho. O estudo concentrou-se no uso da informação por profissionais

⁴ Constance C. Gould. *Information needs in the humanities: na assessment*. Mountain View, Ca.: Research Libraries Group, p. 51; Constance C. Gould and Mark-Handler. *Information needs in the social sciences: na assessment*. Mountain View, Ca.: Research Libraries Group, 1989; Constance C. Gould. *Information needs in the sciences: na assessment*. Mountain View, Ca.: Research Libraries Group, 1991.



dos níveis júnior, intermediário e sênior nas disciplinas de Química, Genética e Ciência da Computação, em instituições acadêmicas privadas e governamentais. Baseado em contagens diárias dos chamados encontros de informação, definidos como “qualquer utilização ou aquisição de informação, idéias ou dados que se relacionam à substância de sua profissão”, o estudo do *Faxon Institute* contém grande volume de informações valiosas e reveladoras sobre a “capacidade de obter informações” percebida nos profissionais investigados, o tipo de fontes que eles usam para obter informação, a frequência com a qual utilizam tais fontes, o tempo que gastam coletando informação, o uso que é dado à informação e, especialmente, seus métodos ou maneiras de obter as informações necessárias.⁵

A **Tabela 1** resume alguns dos resultados do estudo do *Faxon Institute* e indica claramente que os pesquisadores da área científica dependem intensamente, para seus trabalhos produtivos, das fontes de informação prontamente disponíveis. A opinião dos participantes do estudo era de que a biblioteca é tão útil quanto consultar suas próprias bibliotecas particulares ou discutir pessoalmente o assunto. Contudo, eles usavam a biblioteca com muito menos frequência do que as outras formas de obter informações, presumivelmente porque consome-se algum tempo ao se deslocar até a biblioteca e utilizá-la. Além disso, a consulta ao material disponível na biblioteca pessoal de um dado pesquisador ocorreu em um número bem maior de ‘encontros de informação’ do que qualquer outro método e foi, junto com o correio eletrônico, a forma utilizada com maior frequência. E mais, se todos os métodos de consulta a colegas — conversa pessoal, telefonemas, comunicações por escrito, correio eletrônico e fax — forem combinados, a pesquisa do Instituto Faxon sugere que a rede pessoal de colegas de um dado pesquisador ocupa o primeiro lugar e que a biblioteca ocupa um distante terceiro lugar na porcentagem de utilização para solucionar um problema de coleta de informação.

Tabela 1. Utilização dos modos de aquisição de informação.⁶

Modo de informação	Porcentagem do total de encontros de informação em que o modo foi usado	Número médio de vezes utilizado na semana transcorrida	Utilidade percebida modo (classificação média em escala de 10 pontos)
Biblioteca/Arquivo Pessoal	45,0%	9,7	7,9
Biblioteca	31,0%	2,8	7,9
Conversa Pessoal	29,0%	7,7	7,9
Telefone	15,0%	9,5	7,2
Comunicação por escrito	7,0%	4,4	5,9
Base de dados <i>online</i>	6,0%	2,6	6,8
Correio eletrônico	5,0%	14,7	6,2
Fax	4,0%	2,6	6,8
CD ROM	2,0%	1,6	5,1
BBS	0,3%	3,6	4,3

⁵ Eric Almquist. An examination of work-related information acquisition and usage among scientific, technical and medical fields, resultados de um estudo comissionado pelo Faxon Institute for Advanced Studies in Scholarly and Scientific Communications, presented at the 1991 Faxon Institute Annual Conference, Creating users pathways to electronic information, Apr. 1991, Reston, Virginia, exhibit 2.

⁶ Idem, *ibidem*, exhibits 9, 22 e 25.

A biblioteca acessível

Os resultados dos estudos do *Research Libraries Group* e do *Faxon Institute*, em geral, reforçam as conclusões de quase duas décadas de pesquisa sobre o comportamento dos estudiosos na busca da informação.⁷ Existem ainda muitos campos para trabalhos adicionais que venham a refinar e testar rigorosamente estas conclusões, com relação às disciplinas e aos grupos de disciplinas. Precisa-se, em particular, de informações mais detalhadas, no contexto destes estudos, sobre as maneiras como os pesquisadores das várias disciplinas realmente utilizam as bibliotecas. Porém, conforme estes estudos sugerem, se os pesquisadores valorizam extremamente e tendem a optar pelo uso da informação que se encontra prontamente à mão, então uma medida decisiva do sucesso para as bibliotecas que planejam seu curso no futuro será a rapidez com que elas irão disponibilizar a informação para seus usuários.

Um modelo de biblioteca de acesso orientado pode ser esquematicamente retratado conforme mostra a **Figura 1**. Este modelo é, obviamente, uma simplificação. Ele não representa adequadamente todas as características da biblioteca e certamente não constitui a única maneira de caracterizar o sistema. Para fins de discussão, contudo, ele destaca os principais componentes funcionais e suas inter-relações.

Neste modelo, coleções e serviços são os dois componentes primários da biblioteca. A administração da biblioteca é o terceiro elemento principal. A administração apoia e fortalece as funções dos serviços de acesso e de coleções através da organização e execução de serviços dentro da biblioteca, como, por exemplo, gerenciamento de instalações, orçamento, pessoal e tecnologia, além de representar a biblioteca junto aos seus usuários e a outras agências externas. Cabe, também, à administração a incumbência de formular e articular a missão da biblioteca no contexto institucional local.

A missão da biblioteca de acesso orientado é a de gerar, preservar e melhorar o acesso às coleções de conhecimentos registrados. Esta missão governa a relação fundamental entre os serviços de acesso e as coleções da biblioteca e seus usuários, e são de dois tipos:

Acesso intelectual e físico

Os serviços de acesso intelectual, como a catalogação, descrevem e organizam para os usuários da biblioteca a informação em coleções de interesse

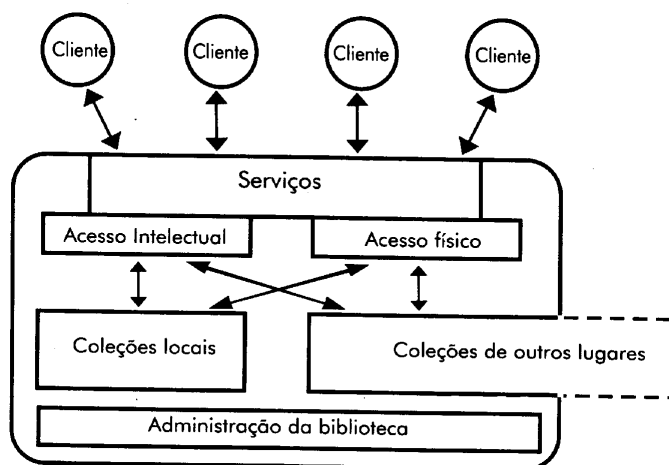


Figura 1. Modelo de biblioteca de acesso orientado

⁷ Ver exemplo, F. W. Lancaster. *The measurement and evaluation of library services*. Washington, D. C.: Information Resources Press, 1977. p. 313. Lancaster faz referência a uma série de estudos que defendem o argumento de que a "seleção de uma fonte de informação se baseia quase que exclusivamente na acessibilidade, sendo a fonte (canal) mais acessível a primeira a ser escolhida; considerações a respeito de qualidade e confiabilidade são secundárias, apesar de serem fatores importantes para influenciar o grau em que o usuário está disposto a aceitar a informação fornecida por uma fonte específica".



local e geral. As descrições de uma obra geralmente se referem, de forma padronizada, a características tais como autor, título, local e data de publicação, editor e formato; a organização é tipicamente por assunto. A disposição de materiais na coleção por assunto faz com que uma série de materiais relacionados se encontre disponível para os usuários da biblioteca e é um serviço de acesso intelectual especialmente valorizado nos estudos em áreas humanas e nos estudos interdisciplinares, onde os métodos associativos fazem dos resultados ‘fortuitos’ de se consultar diretamente as estantes da biblioteca, um aspecto fundamental da pesquisa.⁸ Outras dimensões do acesso intelectual incluem a disponibilidade de uma estrutura de conteúdo do material, frequentemente representada por um índice, e a disponibilidade de um índice remissivo do texto completo do material. Os serviços de acesso físico, como a circulação e os empréstimos interbibliotecas, referem-se aos métodos de distribuição que as bibliotecas utilizam para disponibilizar informações, pertencentes às coleções locais e/ou distantes, diretamente aos seus usuários.

Coleção e renovação da coleção

A noção de coleções de biblioteca refere-se à organização, seleção e meios de armazenagem da informação. Para os usuários de uma biblioteca, as coleções mantidas no local são as mais importantes. Obviamente, as coleções de qualquer biblioteca de pesquisa representam apenas uma pequena fração do conteúdo de informação. Portanto, a responsabilidade por uma coleção local inclui, necessariamente, a responsabilidade de distingui-la e relacioná-la a coleções existentes em outros locais.

Dentro da universidade, as coleções da biblioteca representam um grande, talvez o maior, investimento de capital. Como todo capital, a coleção de uma biblioteca está sujeita a processos distintos e constantes de depreciação e renovação (**Figura 2**): o material é perdido ou extraviado, devendo ser substituído; novas aquisições aumentam a coleção; as coleções são depuradas e parte do material é descartado; o acervo existente se deteriora e deve ser restaurado ou conservado, podendo ser preservado por cópia ou por conversão a outro formato, por exemplo, de papel para microfilme.

Em uma biblioteca de acesso orientado, cada um dos vários processos de renovação da coleção é objeto de uma rigorosa avaliação quanto à missão de serviço da organização: qual a qualidade de geração, preservação ou melhoria do acesso físico e intelectual à coleção? Para os fins desta discussão abordaremos um dos processos de renovação, a saber, a microfilmagem para preservação.

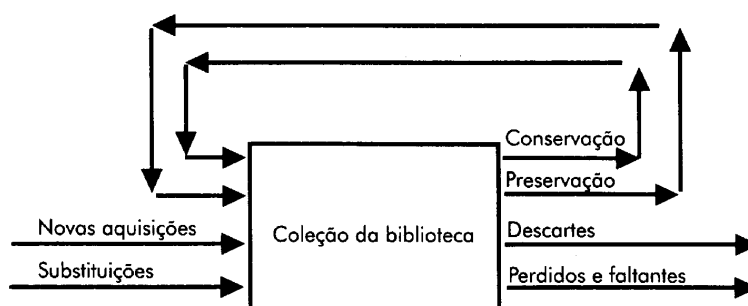


Figura 2. Processo de renovação de coleção

⁸ Lawrence Downler. Conference on research trends and library resources. *Harvard Library Bulletin*, v. 1, no. 2, p. 8, Summer 1990.

Preservando o acesso através da preservação

A deterioração de um documento devido à idade, à acidez ou à utilização excessiva limita tanto o acesso físico quanto o acesso intelectual ao mesmo. Quando um item deteriorado é selecionado para preservação, ele é preparado de forma típica para a microfilmagem (**Figura 3**). Sendo um processo fotográfico, o microfilme reproduz fielmente o material impresso original, incluindo manchas, descolorações, tinta esmaecida, notas dos usuários e as bordas viradas nos cantos das páginas. O processo gera um filme de 35 mm em preto e branco, de alto contraste, em um determinado grau de redução, dependendo do tamanho do original. Os limites práticos da utilização de um meio preto e branco de alto-contraste significam, obviamente, que a microfilmagem não pode reproduzir adequadamente materiais com conteúdo de cor, ilustrações com gradações de cinza ou detalhes impressos extremamente delicados. Para evitar a perda de informação hoje, em função da tecnologia disponível, documentos com conteúdo desse tipo devem ser, sempre que possível, conservados em sua forma original.⁹

Após a reprodução, uma cópia matriz do microfilme é guardada fora da biblioteca, em um ambiente controlado, onde, com base em algumas estimativas, ela durará até 500 anos sem experimentar uma deterioração significativa. Uma cópia positiva e uma negativa do filme são armazenadas na coleção de microfilmes da biblioteca. O filme pode ser posteriormente duplicado a partir da cópia negativa, com relativa facilidade, a um custo de aproximadamente US\$ 20 por rolo e o processo *copyflo* pode gerar uma cópia sobre papel a um custo de cerca de US\$ 0,15 por página. Para itens com uso elevado, a biblioteca pode decidir por disponibilizar uma cópia em papel para leitura. Os usuários da biblioteca podem obter, por meio de solicitação, uma cópia completa do documento em papel ou podem imprimir páginas selecionadas em um copiador de microfilme.

Confrontada com a perda total de conteúdo intelectual de um documento em deterioração, a biblioteca obtém vantagens significativas com a microfilmagem para preservação. O conteúdo é salvo e mantido na coleção, em uma forma compacta e durável, por meio de uma tecnologia que é relativamente simples e bem estabelecida. Mas, por outro lado, o acesso intelectual e físico ao material, por parte dos usuários, é significativamente diminuído.

Diferentemente de um livro, que pode ser carregado e utilizado praticamente em qualquer lugar, o microfilme obriga o usuário a utilizar um equipamento especial de projeção, em um local específico. O usuário pode ficar sabendo do conteúdo intelectual de um documento em microforma, em parte por meio do registro bibliográfico criado durante o processo de microfilmagem, mas, ao converter de papel para microfilme, o processo divide uma coleção, de modo a não ser mais tão fácil beneficiar-se intelectualmente da associação física de um volume na estante com outros de conteúdo similar. Além disso, o microfilme é difícil de ser ‘folheado’ e de ser lido. Com uma leitora de microfilme, o usuário não consegue utilizar, com a mesma facilidade do que em uma cópia em papel, o aparato estrutural interno do volume — índice de assuntos, sumários dos capítulos, prefácios, notas de pé de página, índices remissivos etc. — para obter acesso rápido e eficiente ao conteúdo intelectual do volume, seja ele uma nota, página, capítulo ou alguma combinação dos mesmos¹⁰. Por último, cópias impressas a

⁹ Michael Lesk, op. cit., p. 303.



partir de microfilme, utilizando tanto a tecnologia *copyflo* quanto a tecnologia de copiador, têm recebido poucos elogios quanto às suas características de clareza e de resoluções.

Não é de se surpreender que, dadas estas deficiências, os usuários reclamem intensamente a respeito da utilização do microfilme. Sob um aspecto, essas queixas a respeito das deficiências do microfilme parecem injustas: sem a microfilmagem para preservação, porções substanciais de coleções de bibliotecas estariam completamente perdidas. Neste contexto, a inconveniência de se usar o microfilme parece desprezível. Contudo, dado o valor da acessibilidade à informação para os pesquisadores, a biblioteca é obrigada a assegurar que não pode aplicar nenhuma alternativa tecnológica melhor e economicamente efetiva, a não ser a microfilmagem, para salvar a conteúdo intelectual de materiais em rápida deterioração e, ao mesmo tempo, preservar ou melhorar o acesso dos pesquisadores ao material. Por esta razão, bibliotecas como a de Yale, em que mais de 80% da coleção encontra-se atualmente ou é potencialmente quebradiça, devem investigar os meios, os custos e os benefícios de se acrescentar a seu *kit* de alternativas para preservação novas e promissoras ferramentas, incluindo a tecnologia de imagem digital.

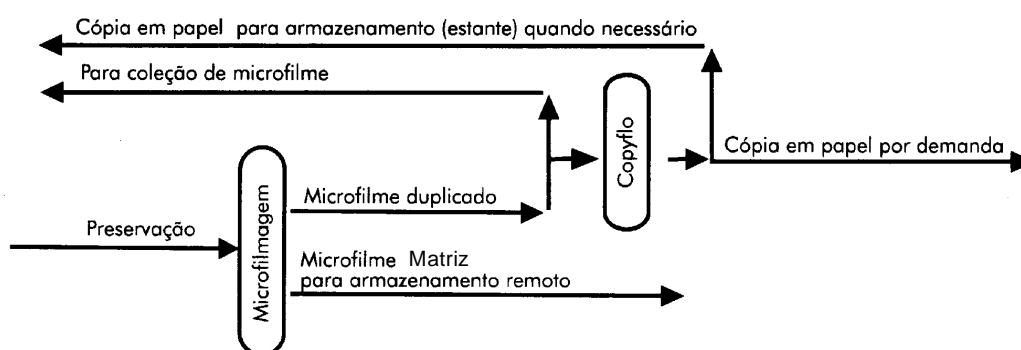


Figura 3. Microfilmagem de preservação

Imagem digital

A tecnologia de imagem digital propicia os meios de se codificar digitalmente documentos escaneados em forma de imagem para armazenagem, transmissão e recuperação em sistemas computadorizados. As imagens digitais produzidas que utilizam essa tecnologia contêm texto, mas o texto não é convertido e não é acessível na forma alfanumérica. Existem recursos para tal conversão, total ou parcial, pela aplicação às imagens digitais, de tecnologias de reconhecimento de caracteres. O

¹⁰ *Blipping*, isto é, a colocação de marcações especiais em fotogramas selecionados, pode facilitar a utilização do microfilme especialmente se as marcações forem codificadas e indexadas em um sistema para recuperação automatizada. A técnica, contudo, tem sido aplicada principalmente em filmes de 16 mm. A UMI estabeleceu normas de uso interno para o *blipping* em filmes de 35 mm. Se essas normas se tornarem amplamente disponíveis e aplicadas de forma generalizada ao microfilme de 35 mm para preservação em bibliotecas, é o que se verá no futuro.

uso da imagem digital aqui idealizado é planejado para antecipar a aplicação eventual do reconhecimento de caracteres, mas não o incorpora no presente momento¹¹.

A digitalização promete melhorar bastante a acessibilidade de materiais de preservação, em comparação com o microfilme (Figura 4). O acesso físico à biblioteca de imagens digitais, e logicamente ao microfilme, requer equipamento especial. Em comparação às leitoras de microfilme, contudo, o equipamento de imagem digital — um poderoso computador com um monitor de resolução relativamente alta — é tecnologia de uso geral onde o acesso à biblioteca digital deve ser apenas uma das aplicações. Além disso, o pesquisador que se encontra fora da biblioteca, e possivelmente até mesmo ausente do campus da biblioteca que possui as imagens, pode ter acesso a elas de maneira relativamente fácil e rápida, uma vez que as imagens digitais podem ser transmitidas pelas redes de alta velocidade já instaladas e interconectadas em muitos *campi* universitários. O pesquisador tem, também, a oportunidade, utilizando a tecnologia de imagem digital, de adicionar integral ou parcialmente o documento de imagem a sua biblioteca pessoal através da requisição de uma cópia impressa. Esta terá, provavelmente, melhor qualidade e resolução mais alta que a equivalente produzida a partir de microfilme e pode até ser substancialmente mais barata, novamente porque a tecnologia subjacente serve para uso geral.

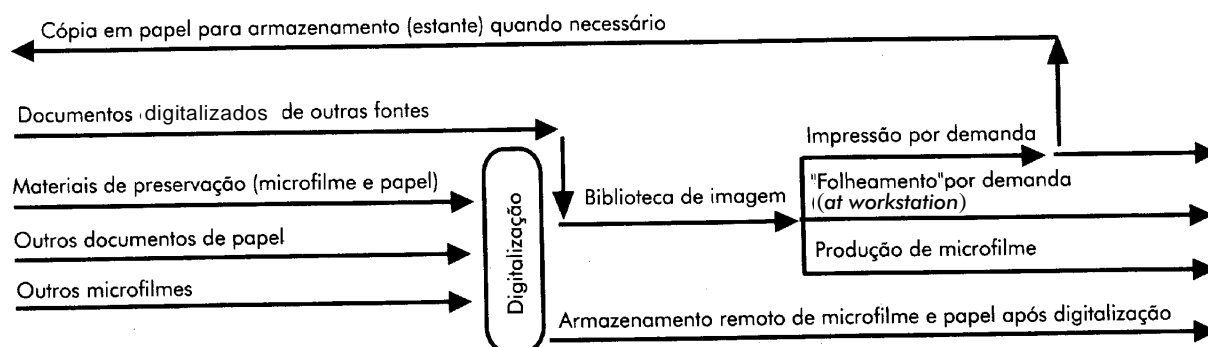
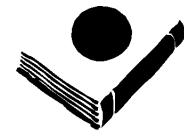


Figura 4. Imagem digital na biblioteca

¹¹ O reconhecimento de caracteres a partir de imagens escaneadas digitalmente levanta uma série desafiadora de questões. Em primeiro lugar, apesar do reconhecimento dos caracteres alfanuméricos estar sendo constantemente aprimorado a precisão da tecnologia atualmente disponível é frustrante. Em segundo lugar, há muito o que fazer para melhorar a habilidade do reconhecimento de caracteres alfanuméricos ao lidar com a grande variedade de tipografias e línguas que são encontradas nos documentos preservados em bibliotecas. Em terceiro lugar, dado o reconhecimento bem sucedido dos caracteres alfanuméricos que compõem uma imagem de texto, é necessário, para propósitos de armazenagem e recuperação, que se relacione os caracteres à estrutura e ao *layout* do documento original, incluindo quaisquer ilustrações, desenhos, diagramas ou outras imagens gráficas que estejam nele contidas.

A maneira como essas questões foram, enfim, abordadas e reconhecidas afetará bastante a acessibilidade dos documentos armazenados na forma eletrônica. Nós pretendemos estar conscientes e informados a respeito das implicações das decisões e escolhas que faremos neste projeto para a aplicação futura da tecnologia de reconhecimento de caracteres. De fato, pretendemos tomar todas as precauções razoáveis para assegurar o sucesso de uma aplicação dessa natureza às imagens digitais que produziremos. Contudo, dada a complexidade e relativa imaturidade da tecnologia de reconhecimento de caracteres, nos referimos à sua aplicação como afim, porém essencialmente fora do alcance do presente projeto.

Para uma discussão frutífera de alguns dos temas associados com a aplicação do reconhecimento de caracteres a documentos de bibliotecas, ver Stuart Wibel, John Handley; Charles Huff. Automated document architecture processing and tagging. In: Donald L. Blumberg, Carol L. Dowling; Claudia V. Weston (Ed.). *Proceedings of the Conference on Application of Scanning Methodologies in Libraries*. Beltsville, Maryland: National Agricultural Library, 1989. p. 3-14.



A imagem digital também promete uma melhoria substancial ao acesso intelectual, em comparação ao microfilme. A capacidade dessa tecnologia de disponibilizar cópias em papel do documento original, de alta qualidade e relativamente baratas, para a biblioteca pessoal dos pesquisadores torna o aparato intelectual padrão — sumário, índice e assim por diante — diretamente acessível. Além disso, diferentemente da prática comum com imagens sobre microfilme de 35 mm, as imagens digitais podem ser indexadas pelos números das páginas e, se os recursos disponíveis permitirem, por estruturas tais como seção, capítulo e fascículo. Com um índice assim e um *software* para utilizá-lo, os pesquisadores podem percorrer rapidamente o documento digital e ‘folhear’ seu conteúdo intelectual de forma bem mais eficiente do que a permitida por documentos preservados em microfilme.

Nas formas aqui descritas, a tecnologia de imagem digital não proporciona uma melhoria substancial, em relação ao microfilme, quanto a proporcionar um substituto eficiente para a consulta de volumes em estantes. Contudo, estudos recentes sugerem que a adição de informações suplementares, como transcrições de índices, às descrições bibliográficas disponíveis *online*, irá permitir uma oportunidade melhor para efetuar associações intelectuais, do que com métodos tradicionais de classificação e a disposição dos documentos em papel nas estantes da biblioteca por assunto¹². À medida em que o conteúdo dos documentos é adicionado aos registros bibliográficos *online* e os benefícios resultantes da melhoria do acesso intelectual aos materiais, em todos os formatos, se tornam aparentes, torna-se possível alcançar uma substancial melhoria no aproveitamento do material no formato de imagem digital por meio da criação de conexões eletrônicas diretas entre a informação suplementar do registro bibliográfico e as imagens digitais do documento a que o registro se refere. A criação de tais conexões exigirá um trabalho considerável, particularmente no sistema bibliográfico *online*, mas a visão de suas possibilidades significa também que se deve planejar, de forma cuidadosa e criativa, os índices da paginação e da estrutura do documento dentro dos sistema de imagem digital, para que as imagens digitais possam ser relacionadas, de maneira fácil e confiável, à informação do conteúdo armazenada em outras formas legíveis por computador, como o registro bibliográfico *online*.

Nossa visão da biblioteca do futuro está relacionada às diferenças e às interrelações principais que identificamos entre

- as coleções da biblioteca que incluem os modos — de armazenagem de informações e do conhecimento registrado — selecionados;
- o acesso físico, que inclui os modos de distribuição das coleções; e
- o acesso intelectual, que inclui os modos principais de uso das coleções.

Assim como esperamos que a tecnologia da imagem digital introduza maior flexibilidade e melhore os modos de utilização e distribuição de coleções, particularmente em comparação com o microfilme, esperamos, também, que essa tecnologia produza efeitos similares nos modos de armazenagem na coleção. Neste contexto, é crucial distinguir entre armazenagem arquivística e armazenagem para uso e distribuição.

¹² Ver, por exemplo, Karen Markey. *Subject searching in library catalogs: before and after the introduction of online catalogs*. Dublin, Ohio: Online Computer Library Center, p. 75-117, 1984; Richard Van Orden. Context-enriched access to electronic information: summaries of selected research. *Library Hi Tech*, v. 8, no. 3, p. 27-32, 1990; e Martin Dillon e Patrick Wenzel. Retrieval effectiveness of enhanced bibliographic records. *Library Hi Tech*, v. 8, no. 3, p. 43-46, 1990.

Na biblioteca do futuro, presumimos que o microfilme continuará a ter um papel importante, talvez ainda permanecendo como o meio preferido de armazenagem de uma versão arquivística dos documentos preservados¹³. Para a utilização e a distribuição dos materiais preservados, contudo, as imagens digitais se tornarão provavelmente o meio de armazenagem preferido. Dependendo dos custos, a microfilmagem poderá permanecer como uma etapa preliminar em um processo de preservação que eventualmente levaria à digitalização, ou então, o filme poderá ser gerado a partir da imagem digital. Em ambos os casos, esperamos que a tecnologia digital aproveite a base substancial de documentos já preservados em microfilme e atualmente disponíveis apenas nesta forma.

Finalmente, nossa visão da utilização da tecnologia de imagem digital sugere alguns efeitos mais amplos. Apesar de esperarmos que a biblioteca digital seja estabelecida a partir de um impulso inicial visando à preservação e à melhoria do acesso à enorme quantidade de livros quebradiços e em deterioração nas estantes das bibliotecas, nós também esperamos um enriquecimento das bibliotecas ao longo do tempo, com documentos na forma de imagem digital que tenham sido criados por várias outras razões, tanto dentro quanto fora da biblioteca. Além disso, a tecnologia de imagem digital terá provavelmente um efeito significativo, talvez profundo, na maneira segundo a qual a biblioteca do futuro encarará, organizará e valorizará seu espaço e sua equipe de funcionários. Por exemplo, na medida em que a biblioteca digital surja como uma forma de renovar e substituir grandes estoques de materiais de papel e de microfilme, e na medida em que a biblioteca digital seja armazenada e acessível remotamente da biblioteca propriamente dita, o uso do espaço central de estanteria, onde se armazenam documentos na biblioteca, quase que certamente será modificado e o equilíbrio das funções da equipe de funcionários — tradicionalmente associadas à circulação e à recolocação nas prateleiras de coleções armazenadas em papel — deverá ser alterado em função dos modos de distribuição de documentos gerados sob demanda a partir de formatos eletrônicos¹⁴.

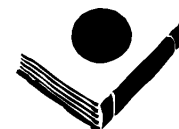
Uma economia de escolhas

As bibliotecas do futuro não alcançarão, em uma única e rápida etapa, a visão aqui apresentada com relação ao uso da tecnologia digital na preservação e na melhoria do acesso ao conhecimento registrado. Contudo, as promessas de benefícios em serviços e de uma possível economia de custo mobilizará, de várias formas, os bibliotecários na direção dessa tecnologia. Restrições orçamentárias e outros impedimentos organizacionais irão forçá-los, ainda, a investir nessa tecnologia de outras formas. Se a promessa justificar o investimento, a maioria irá incorporar a tecnologia de reprodução digital de imagem da mesma maneira como adotou e assimilou outras tecnologias no passado: através de uma série de decisões específicas e escolhas regidas pela função e pelas necessidades de suas instituições.

Aqueles que desenvolvem a tecnologia de imagem digital para utilização na preservação de bibliotecas devem ter em mente o princípio de aperfeiçoar seu poder e sua influência na mudança

¹³ Sobre a contínua utilidade do microfilme na indústria, veja Whitney S. Minkler. *Optical disks vs. Micrographics: is it an adversarial problem?* *Micrographics and Optical Technology*, v. 7, no. 4, p. 141-149, 1989; e John Blake. *War over optical disk or microfilm ends, future contains multiple media.* *Micrographics and Optical Technology*, v. 8, no. 3, p. 141-143, 1990.

¹⁴ Ver, por exemplo, Clifford A. Lynch e Edwin B. Brownrigg. *Library applications of electronic imaging technology.* *Information Technology and Libraries*, v. 5, no. 2, p. 100-102, June 1986.



organizacional. A economia no gerenciamento e administração dos recursos de uma biblioteca, como na maioria das grandes instituições, se reflete na racionalização de escolhas, mesmo quando se trata de uma tecnologia potencialmente revolucionária. O desenvolvimento da aplicação da imagem digital deve, então, produzir resultados no tocante aos meios, custos e benefícios que facilitem e informem, em vez de distorcer e enganar, sobre tal economia.

Esta seção almeja identificar um conjunto de escolhas funcionais cruciais que, provavelmente, governarão a incorporação da tecnologia de imagem digital nos processos de preservação de bibliotecas. A próxima, delinea muitas das exigências técnicas detalhadas para a arquitetura de um sistema que possa acomodar, de maneira flexível, as diferentes escolhas que as bibliotecas provavelmente farão. A seção final apresenta um plano de trabalho para o desenvolvimento da arquitetura e para a geração da informação que as bibliotecas necessitarão ao fazerem uma escolha cautelosa referente aos custos e benefícios dessa tecnologia de imagem.

O foco na conversão a partir do microfilme

A Universidade de Cornell, uma pioneira na utilização da imagem digital para fins de preservação em biblioteca, optou pela investigação de métodos para superar as limitações da microfilmagem para preservação através do escaneamento de documentos em deterioração diretamente para a forma de imagem, em vez de microfilmá-los. Trabalhos adicionais fazem-se necessários visando expandir as possibilidades de alimentação direta, incluindo métodos para o manuseio de documentos de grandes dimensões, documentos coloridos e outros tipos de documentos especiais. Além disso, como podem ser eletronicamente realçadas pela eliminação de manchas e outras marcas e pela acentuação do contraste de tintas esmaecidas, as imagens digitais deverão se tornar, em última análise, uma fonte de qualidade superior à fonte original em papel para a cópia sobre microfilme. Os meios para a geração de uma cópia sobre microfilme a partir da imagem digital devem, pois, ser sistematicamente explorados.

O papel, contudo, não é a única fonte de alimentação possível para a criação de uma biblioteca digital de materiais preservados. A biblioteca da Universidade de Yale busca demonstrar à grande comunidade bibliotecária a viabilidade de uma fonte alternativa, através do desenvolvimento de meios e da identificação dos custos e benefícios de se escanear materiais preservados a partir da microfilmagem. Embora seja alimentado por microfilme em vez de papel, o sistema de reprodução de imagens desenvolvido no projeto da Universidade de Yale também proporcionará as mesmas capacidades funcionais básicas que o projeto em desenvolvimento na Universidade Cornell. No sistema completo, os usuários da biblioteca poderão ‘folhear’ a biblioteca de imagens a partir de um terminal de computador (**Figura 5**). Eles poderão, também, obter uma versão impressa do documento digital, sempre que requisitado; a própria biblioteca poderá gerar uma cópia sobre o papel do documento preservado e colocá-la nas estantes da biblioteca. Além disso, a biblioteca de imagens será armazenada separadamente da coleção nas estantes. Devido à presumida acessibilidade da biblioteca digital, tanto de dentro quanto de fora da biblioteca, a versão em microfilme dos documentos digitalizados pode ser idealmente removida da biblioteca para um depósito mais remoto e barato.

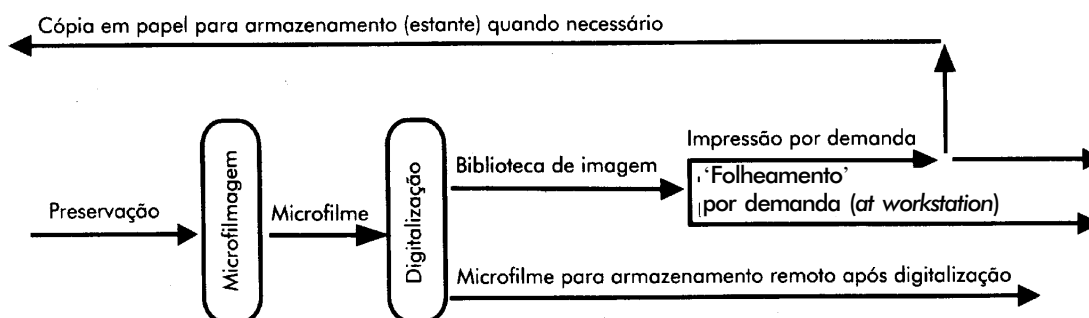


Figura 5. Conservação de microfilme de preservação para imagem digital

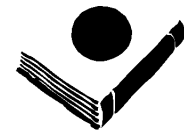
O sistema da Universidade de Cornell está sendo desenvolvido, inicialmente, para manejar uma biblioteca de imagens com 1.000 volumes. Além do desenvolvimento de uma fonte alternativa de entrada, o projeto da Universidade de Yale testará a possibilidade de implementar projetos como o da Universidade de Cornell, em grande escala, aumentando o acervo da biblioteca digital, em uma ordem de magnitude, para 10 mil volumes. O tamanho da biblioteca digital de Yale, comparado à de Cornell, ajudará a demonstrar as economias do sistema quando implementado em escala e a assegurar à comunidade de bibliotecários que é possível medir realisticamente os custos associados à adição de componentes ao sistema e à manutenção destes componentes ao longo do tempo.

Um modelo de investimento em incrementos

Os custos e benefícios de um sistema de reprodução de imagens dependem largamente, embora não exclusivamente, dos principais componentes funcionais do sistema, do método e do ritmo de sua implementação. Não é necessária a adoção de todos os componentes possíveis para a criação de um sistema de trabalho. Considerando a missão a ser cumprida pelo sistema, alguns componentes são essenciais e devem ser implementados, já outros são menos importantes, podendo ser instalados opcionalmente ou posteriormente à base de incrementos. A análise que se segue distingue os principais componentes funcionais de um sistema de imagem digital para preservação, ordena os componentes em relação ao objetivo do sistema e os arranja em uma seqüência cumulativa de etapas. A seqüência sugere uma estratégia plausível, oferecendo opções de escolha a cada nível para uma biblioteca que esteja investindo na tecnologia de imagem digital. Para dar suporte às opções, o desenvolvimento do projeto de reprodução de imagens em Yale precisa ser planejado de modo a permitir a medição dos custos e benefícios referentes ao incremento de componentes do sistema, identificados em cada etapa da seqüência.

O sistema básico

Se o objetivo da preservação é de salvaguardar e otimizar o acesso físico e intelectual à documentação impressa em deterioração, e considerando-se que o material já foi microfilmado, o sistema de imagem digital mais simples e direto é aquele que produz, a partir do microfilme, uma cópia impressa, de alta qualidade, do documento original para a biblioteca colocar em suas prateleiras. Os componentes essenciais de um sistema desse tipo seriam:



- um escaner para microfilme;
- um dispositivo de armazenagem temporária para guardar as imagens digitalizadas;
- uma estação de controle de qualidade para revisar as imagens, realçá-las sempre que possível e escanear o filme novamente quando necessário;
- uma rede de alta velocidade para a transferência das imagens ao serviço de impressão;
- um serviço de impressão, para colocar em fila e também controlar o trabalho de impressão; e
- uma impressora digital de alta qualidade.

Uma empresa de serviços poderia, naturalmente, se encarregar parcial ou totalmente dessas funções. Poderia escanear com controle de qualidade e imprimir o documento; poderia escanear com controle de qualidade e devolver à biblioteca o conjunto de imagens digitalizadas para impressão. Será importante, para o projeto demonstrativo de Yale, acumular experiência interna suficiente em digitalização, de modo que se possa transmitir à empresa de serviços os padrões e as expectativas para conversão e indexação, para comparar integralmente os custos de um serviço interno com os custos da contratação de uma empresa de serviços que realiza as mesmas funções.

Devido ao fato de que o processo *copyflo* já é capaz de produzir cópias impressas diretamente a partir do microfilme, uma biblioteca provavelmente achará este sistema básico de reprodução de imagens atraente se, e somente se, os custos de digitalização e impressão forem inferiores aos custos do *copyflo* ou se a cópia gerada pela impressora digital for de qualidade superior. Considerando-se como provavelmente o é, que a resolução da imagem digital gerada a partir do microfilme é da ordem de 300 pontos por polegada (dpi), então a qualidade de impressão certamente será igual ou superior à do processo *copyflo*. Michael Lesk estimou, ainda, que o custo de uma conversão simples do microfilme para imagem digital é de aproximadamente 2 centavos de dólar norte-americano por fotograma¹⁵. Estimativas anteriores de Cornell sugerem que os custos de impressão de um documento a partir dessas imagens são aproximadamente de 10 dólares por livro de 300 páginas (sem encadernação) ou de 3,3 centavos de dólar norte-americano por página.

Estas estimativas incluem pressupostos acerca de custos da mão de obra, produtividade, volume e métodos de financiamento e amortização de equipamentos que podem ou não valer para qualquer circunstância. Contudo, reconhecendo as limitações de tais estimativas, compare-as ainda ao custo atual de 15 centavos de dólar norte-americano por página (sem encadernação) do processo *copyflo*. Um sistema de imagem digital restrito aos seus componentes essenciais e simplesmente capaz de digitalizar e imprimir imagens de microfilme irá não somente produzir cópias de qualidade superior, mas também proporcionar mais de 60% em economia com relação aos meios presentes de impressão a partir do microfilme. Obviamente, com componentes adicionais, um sistema de imagem digital pode desempenhar uma função bem maior do que a do sistema aqui descrito, mas os riscos do investimento aumentam proporcionalmente.

15 Michael Lesk, op. cit., p. 307.

Impressão sob demanda

Se um sistema de imagem digital básico habilita a biblioteca a restringir o acesso a um documento em deterioração através da geração de uma cópia de fac-símile de alta qualidade e a um custo relativamente baixo, então pode-se legitimamente imaginar qual o investimento adicional que seria necessário para produzir a cópia impressa sob demanda para um pesquisador colocar em sua biblioteca pessoal. Um sistema de imagem digital capaz de impressão sob demanda necessita da incorporação dos seguintes componentes ao sistema básico:

- um dispositivo de armazenagem permanente para as imagens do documento na forma digital;
- um índice a nível de título para o documento na forma de imagem, de modo que se possa distinguir um documento de outro no meio de armazenagem;
- uma entrada em uma descrição bibliográfica nova ou já existente, presumivelmente registrada no catálogo *online* da biblioteca, que indique ao usuário que o documento se encontra armazenado na forma digital e está disponível para impressão sob demanda;
- um serviço de requisição de impressão; e
- um serviço de imagem que extraia as imagens sob o identificador único e as transmita à impressora.

A otimização da impressão sob demanda sujeita-se também a custos substanciais, particularmente com respeito ao mecanismo de armazenagem. A armazenagem de imagens digitais é tecnicamente complexa e será tratada com maior detalhe na discussão sobre a arquitetura do sistema, na próxima seção. É importante citar aqui, contudo, que a noção de ‘permanência’ na armazenagem digital envolve os custos de se atualizar periodicamente os arquivos de imagem — e relativamente com mais frequência que livros e microfilme — na medida em que os meios de armazenagem, os formatos de arquivo digital e o equipamento utilizado para se ter acesso a esses arquivos sofrem mudanças. É possível, mas ainda não provado, que a economia feita em espaço, a partir de uma armazenagem mais densa e compacta e em outras eficiências adquiridas com a mudança de tecnologia, poderia contrabalançar os custos de atualização de arquivos de imagem e, assim, fazer com que o processo de renovação seja auto-sustentável.

O sistema de indexação neste estágio é simples de ser criado e utilizado. Consiste na aplicação de um número de classificação ou outro identificador único à coleção de imagens do documento digital. O identificador único proporciona a chave que conecta a entrada bibliográfica do catálogo ao título na biblioteca digital, igual a um número de chamada que permite ao usuário dirigir-se ao catálogo, a um livro na estante ou a um rolo de microfilme em uma gaveta específica de arquivo.

Naturalmente, um pesquisador nem sempre pode dizer, a partir da entrada de um catálogo, se um documento é ou não é relevante ao problema ou à pesquisa de seu interesse. Antes de requisitar uma cópia impressa, pode ser necessário ‘folhear’ o documento em microfilme na biblioteca para determinar sua relevância. Os investimentos em um sistema de imagem digital que propicie um serviço de impressão sob demanda incluem, portanto, os custos de armazenagem e indexação dos documentos na forma digital, sem, contudo, prescindir da exigência de que a biblioteca mantenha uma versão dos mesmos em microfilme, para que o usuário possa ‘folheá-los’.



'Folheando' um documento online

Considerando-se a armazenagem permanente necessária ao serviço de impressão sob demanda, uma outra extensão natural dos sistema de imagem digital seria permitir ao usuário 'folhear' um documento *online* diretamente na forma de imagem digital. A capacidade para se 'folhear' *online* requer, adicionalmente àqueles já identificados, os seguintes componentes:

- um terminal de computador e uma impressora para que o usuário possa recuperar, exibir, navegar e imprimir imagens selecionadas do documento; e
- um serviço de apresentação que adapte a apresentação da imagem ao tipo específico de estação de 'folheamento' em uso.

A melhoria destes componentes do sistema de imagem digital proporciona ao usuário capacidades funcionais que, em pelo menos um aspecto, se assemelham ao 'folheamento' de um documento na forma de microfilme. Devido ao fato de que o documento neste estágio está indexado apenas pelo título, com um identificador único, não se pode ir automaticamente até uma página ou seção específica do documento. Como ocorre com uma leitora de microfilme, deve-se mover para frente e para trás, através do documento, imagem por imagem. A principal diferença funcional em relação ao uso do microfilme, nesse estágio, é que o usuário não tem que localizar o rolo de filme manualmente, mas depende do *software* do sistema para encontrar e carregar o documento automaticamente. Além disso, após 'folhear' o documento, o usuário pode solicitar que o sistema gere uma cópia impressa de alta qualidade para uso pessoal.

É dever dos responsáveis pelo desenvolvimento do sistema assegurar que as características de 'folheamento' *online* de impressão sob demanda incorporadas ao sistema, neste estágio de investimento, sejam estimulantes e suficientes ao pesquisador, para que o mesmo as utilize em lugar do microfilme. Partindo do pressuposto de que elas realmente o sejam, a biblioteca poderá começar a acumular pelo menos algumas economias em microfilme da biblioteca. A biblioteca poderá também contemplar melhorias significativas na capacidade de 'folheamento'.

Acesso remoto

A biblioteca pode estender a capacidade de 'folheamento' do sistema de imagem, sob um certo aspecto, tornando-o acessível remotamente. Dentro da biblioteca, a apresentação das funções de 'folheamento' pode ser controlada pela limitação do tipo de estação de trabalho utilizada. O acesso remoto, contudo, requereria um servidor com maior qualificação para suportar a variedade de estações de trabalho e capaz de lidar com as imagens que os pesquisadores utilizam fora da biblioteca. Além do servidor de imagem e da rede local do campus também necessitaria de capacidade para suportar o tráfego mais intenso. Se a função de 'folheamento' pode, ou não, ser estendida para fora da rede do campus, em termos práticos, é uma questão de considerável interesse que exigirá pesquisa adicional e um grande volume de testes.

'Folheamento' no nível de página

A biblioteca pode, também, estender a capacidade de 'folheamento' do sistema de imagem, tornando as imagens de um documento diretamente acessíveis através do número da página. O investimento no 'folheamento' em nível de página depende diretamente dos custos de geração, manutenção e suporte do *software* de 'folheamento' e de um índice de páginas para o documento de

imagem. Não se pode emitir um comando para ir a uma página específica no documento até que, e a não ser que, alguém, deliberadamente, se dedique à criação de um índice conectando as imagens com os números das páginas do documento. As páginas podem apresentar numeração romana ou arábica. Por outro lado, algumas páginas podem até não estar associadas com um número específico em qualquer forma de notação. O sistema de indexação e recuperação necessita refletir, todavia, sobre essas distinções, variações e peculiaridades.

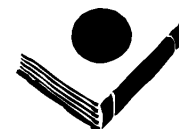
‘Folheado’ em nível de estrutura do documento

A biblioteca pode estender a capacidade de ‘folheamento’ do sistema de imagem ainda mais profundamente, tornando as imagens de um documento diretamente acessíveis através de suas próprias divisões estruturais internas, tais como página de título, sumário, parte, capítulo, seção e índice. Isto é, sem mover-se através do documento página por página, o usuário pode emitir um comando para ir imediata e automaticamente ao terceiro capítulo, à parte três, ou à página do título. Novamente, a escolha para se implementar um sistema de imagem com este nível de funcionalidade depende diretamente do interesse da biblioteca na criação de um índice conectando uma imagem ou um conjunto de imagens a um ou mais elementos estruturais específicos do documento. Uma pequena variação desse sistema proporcionaria ao usuário a capacidade de estruturar temporariamente o documento, durante o processo de ‘folheamento’, com marcadores de página. Uma variação ainda mais ambiciosa seria permitir ao usuário estruturar o documento, completa ou parcialmente, a seu próprio gosto e salvar o índice estrutural resultante, sob seu próprio nome, para uso posterior.

A análise aqui apresentada de um sistema de imagem digital para bibliotecas, que tem a finalidade de preservar e melhorar o acesso a materiais em deterioração e é derivado da conversão de microfilme para formato de imagem, não esgota todos os usos possíveis de um sistema deste tipo. Por exemplo, pode-se decidir pela digitalização de uma coleção de documentos baseando-se no fato de que a coerência intelectual do material justifica o esforço de torná-lo mais acessível. Alternativamente, para auxiliar na diminuição dos custos de armazenagem e indexação, pode-se optar pela digitalização apenas de documentos de uso intenso, como uma etapa intermediária, ou mesmo como uma alternativa à colocação de uma cópia em papel nas estantes. Pode-se, ainda, optar, apenas, pela digitalização de documentos, na medida em que são requisitados para uso, deixando as porções não utilizadas da coleção em microforma.

Esta análise também não esgota todas as características possíveis de um sistema de imagem digital para preservação. Como exemplo, uma vez que a tecnologia de reconhecimento de caracteres esteja suficientemente amadurecida, melhorias adicionais ao sistema de imagem digital tornar-se-ão possíveis. Um documento em forma de imagem poderá ser, parcial ou totalmente, convertido para um arquivo texto de caracteres alfanuméricos e as imagens do documento poderiam, potencialmente, ser conectadas a uma porção do arquivo de texto, como o índice, que é armazenado na descrição bibliográfica do documento de imagens ou a um índice remissivo de palavras-chave gerado a partir do arquivo de texto completo. Avanços deste tipo poderiam enriquecer bastante o valor intelectual da biblioteca de imagens.

A análise feita nesta seção, contudo, destaca as principais dimensões funcionais de um sistema de imagem digital para preservação. Sugere, também, um arranjo plausível de escolhas que deve motivar uma biblioteca a investir economicamente em um sistema deste tipo. Obviamente, deve-se distinguir entre a motivação e a necessidade funcional e a arquitetura técnica subjacente ao sistema de imagem digital — que serve aos propósitos da biblioteca e satisfaz suas necessidades — além do



plano de trabalho para um projeto demonstrativo que desenvolva um sistema completamente articulado dentro da arquitetura e gere informação suficiente sobre o sistema, de forma a permitir às bibliotecas tomar decisões prudentes e práticas no tocante a seus custos e benefícios.

Arquitetura do sistema

A arquitetura — o projeto e a estrutura global — de um sistema para criação, armazenagem, recuperação e impressão de documentos de biblioteca na forma de imagem evoluiu gradualmente durante a última década. Na pesquisa inicial e no desenvolvimento de projetos piloto, especialmente na Biblioteca do Congresso e na Biblioteca Nacional de Medicina norte-americanas, testaram, refinaram e validaram idéias-chave de projeto na medida em que a tecnologia de imagem digital emergiu: as estações de trabalho realizam funções com imagem digital, são distribuídas numa rede de alta velocidade, utilizam os recursos necessários na rede e interagem umas com as outras de acordo com um padrão cliente/servidor¹⁶. Em seu projeto de preservação digital, a Universidade de Cornell incorporou um projeto de estação de trabalho distribuída, no padrão cliente/servidor, em sua arquitetura de sistema; Yale fará o mesmo.

Outros princípios de projeto mais gerais também se aplicam à arquitetura de sistemas de imagem digital. Alguns destes princípios governarão o sistema que vem sendo desenvolvido em Yale e deve-se articulá-los, em vez de assumi-los. Inicialmente, os dados, tanto de imagens quanto de seus índices, constituirão o recurso principal gerado no projeto de Yale. O *software* e o *hardware* necessários para criar, armazenar e utilizar os dados mudarão e serão substituídos. Os dados permanecerão, contudo, e deverão ser facilmente conversíveis de um sistema para outro. Em segundo lugar, baseando-se no argumento da seção anterior sobre a necessidade de escolha, os componentes do sistema devem ser desenvolvidos de forma modular, para poderem ser fácil e prontamente adicionados, eliminados ou alterados quando necessário. Em terceiro lugar, para assegurar flexibilidade na adição, eliminação ou alteração no sistema de imagem digital em Yale, os componentes utilizados devem obedecer às normas oficiais ou industriais, ou ser construídos para interfaces padrão.

Relacionando estes princípios na ordem inversa, esta seção identifica os padrões relevantes para o sistema de imagem digital em Yale, revê os componentes do sistema e os aspectos técnicos a eles associados e considera a exequibilidade global do projeto, evidenciando áreas em que o interesse pelos dados devem garantir atenção especial a características específicas dos sistema.

Normas técnicas

As normas relevantes que se aplicam ao projeto e à construção do sistema de imagem digital em Yale se encaixam em três grupos distintos: normas de comunicações, de armazenagem e de aplicação¹⁷. Assim como a maior parte das principais universidades de pesquisa dos Estados Unidos, a Universidade

¹⁶ Felix P. Krayeski. Transition of na image system: from paper to microfiche to optical disk. Congressional Research Service, The Library of Congress, June 1990; Frank L. Walker; George R. Thoma. Access techniques for document image databases. *Library Trends*, v. 38, no. 4, p. 751-786, Spring 1990; Frank L. Walker. Issues in document conversion. In: Donald L. Blamberg, Carol L. Downing; Claudia V. Weston. (Ed.). *Proceedings of the Conference on Application of Scanning Methodologies in Libraries*. Beltsville, Maryland: National Agricultural Library, p. 45-60, 1989.

¹⁷ Para uma boa visão geral das normas técnicas relevantes, ver Gerry Walter. "Standards help advance document management system progress. *Optical Memory News*. p. 25-27, May, 1990.

de Yale comporta uma variedade de protocolos de rede, mas tem se concentrado nos chamados padrões *ethernet*, incluindo o TCP/IP como protocolo escolhido para comunicações em rede. Componentes do sistema podem operar em uma sub-rede do campus, utilizando um tipo diferente de protocolo padrão, por exemplo, *token ring* (rede em anel) no nível de conexão de dados. Contudo, a sub-rede deve fornecer uma porta confiável e satisfatória para a espinha central da rede, baseada na *ethernet*, para comunicação com outros componentes do sistema.

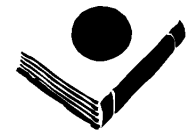
Para armazenagem de dados de imagens, o sistema de Yale suportará o formato TIFF (*Tagged Image File Format*), que proporciona a padronização em nível de cabeçalho, necessária para permitir o intercâmbio de imagens. Devido aos arquivos de imagem serem muito grandes, uma norma técnica se faz também necessária para a compreensão dos arquivos a uma fração de seu tamanho original, para que se possa armazená-la e comunicá-los de forma mais econômica. O padrão de compressão do grupo 4 do CCITT é deficiente em alguns aspectos, porque não inclui páginas cujas dimensões sejam superiores às do tamanho carta, nem manipula bem as imagens com escala mais extensa de gradações de cinza ou coloridas. Grupos de normalização técnica estão cooperando no momento para o desenvolvimento de padrões de compressão mais adequados. Por enquanto, o padrão do grupo 4 do CCITT deve servir às necessidades deste projeto.

Quando se trata das aplicações, a informação bibliográfica, referente a um documento na forma de imagem será incorporada ao catálogo *online* da biblioteca em formato padrão de catalogação legível por computador, MARC - *Machine-readable Cataloging*. As convenções existentes para o MARC não são completamente adequadas à descrição da versão em imagem digital de um documento, mas a respectiva organização de normas técnicas já esboçou uma revisão adequada da norma. A Universidade de Yale adotará uma solução temporária até que a norma revisada seja aceita.

Duas outras normas técnicas são também apropriadas no nível das aplicações. Em primeiro lugar, índices que abrem um documento na forma de imagem para o 'folheamento' nos níveis de título de página e de estrutura do documento, dependem de trabalho que descreva os conteúdos das imagens, defina e torne possível trocar e apresentar um conjunto particular de imagens como um conjunto de entidades de ordem superior (por exemplo, um livro, um capítulo, um conjunto de páginas). Normas técnicas relativas ao conteúdo do documento devem se aplicar (ou ser apropriadamente estendidas para se aplicar) à criação, à armazenagem e ao uso dos vários níveis de índices estruturais. Alguns fornecedores, incluindo a IBM, têm difundido seus próprios padrões internos para conteúdo de documento. Contudo, a norma ISO 8613, intitulada *Information Processing - Text and Office Systems - Office Document Architecture (ODA) and Interchange Format (ODIF)*, é o padrão internacional que prevalece e o projeto de Yale procurará manter-se em conformidade com ela.

Em segundo lugar, os arquivos de estrutura, que descrevem o conteúdo dos documentos de imagem em concordância com a ISO 8613, devem ser armazenados e recuperados de um arquivo de base de dados conectado referencialmente aos arquivos de imagens de documento. Esperamos que a base de dados seja relacional. Esperamos, também, que ela incorpore e esteja em conformidade com o padrão SQL (*Structured Query Language*)¹⁸.

¹⁸ A apresentação de um índice da estrutura do documento em uma forma relacional padronizada, eventualmente facilitará a criação de conexões entre o documento de imagem digital e qualquer informação sobre seu conteúdo armazenada em outras fontes legíveis por computador, tais como a descrição bibliográfica *online* do documento.



Componentes do sistema

Uma representação geral da arquitetura do sistema para o projeto de imagem digital para preservação em Yale é apresentada na **Figura 6**. Os componentes estão todos conectados à rede *ethernet* de alta velocidade do campus, que tem um núcleo de fibra óptica e velocidades atuais de

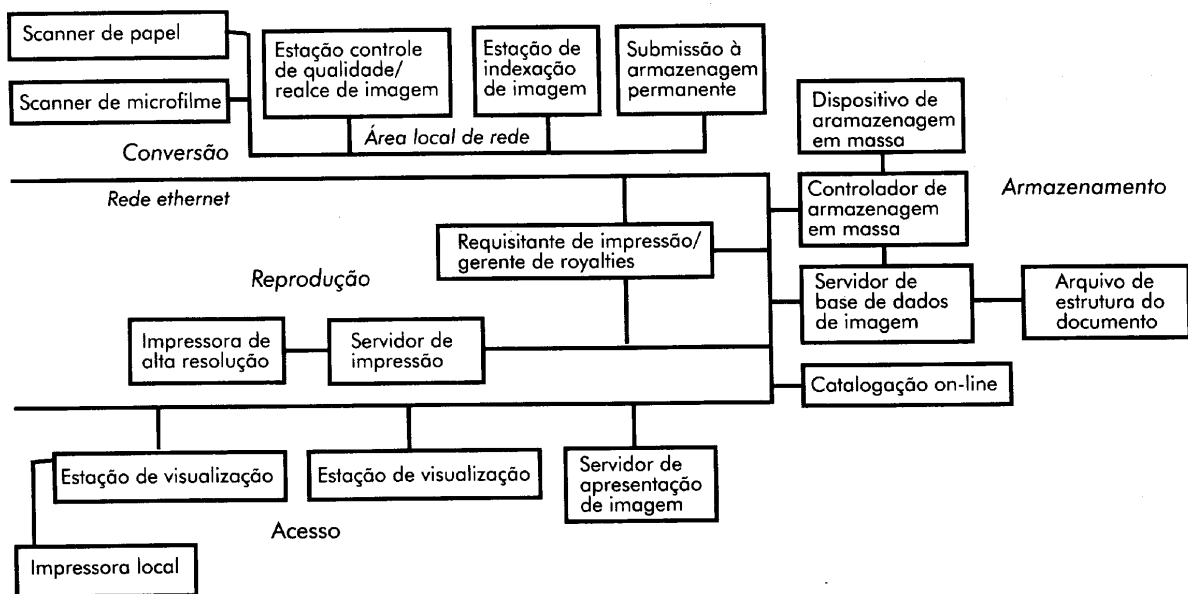


Figura 6. Arquitetura do sistema de imagem

transmissão de até 10 megabytes por segundo. Os componentes chave do projeto são: um subsistema de conversão, um subsistema de reprodução (ou impressão), um subsistema de armazenagem e um subsistema de acesso.

Conversão

O propósito do subsistema de conversão é receber, como alimentação, uma série de imagens de fotogramas de microfilme que constituam um documento (livro, folheto ou periódico) e tirar uma ‘fotografia digital’ de cada quadro, de forma que a imagem seja armazenada como um arquivo de computador e possa ser exibida em um monitor de computador. O processo de conversão assegura a qualidade física da digitalização, que inclui o controle da resolução, bem como do conteúdo de imagem, contraste, distorção e seqüência. O processo também serve para caracterizar, ou indexar, o conjunto de arquivos de imagem de forma que eles retenham sua identidade como um documento único para posterior armazenagem e recuperação. O passo final no processo é submeter o conjunto de arquivos, como uma entidade documental, para armazenagem permanente.

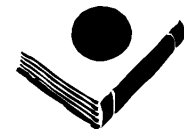
O subsistema de conversão inclui um digitalizador de microfilme e papel e, no mínimo, uma estação de trabalho controladora em que um operador possa realizar as funções de controle de qualidade, indexação e confinamento (armazenagem permanente). Para proporcionar uma maior

flexibilidade, a estação de trabalho e o digitalizador deve estar conectados por uma rede local de alta velocidade. Com a interconexão via rede, novas estações de trabalho podem ser adicionadas para dividir as tarefas, mas cada função pode, ainda assim, ser realizada em todas as estações. Alternativamente — e isto pode ser necessário para maior eficiência —, o digitalizador pode estar diretamente conectado a uma estação de trabalho controladora, na qual um operador realiza as funções de controle de qualidade. Após a conclusão do processo, o operador pode então enviar os arquivos, utilizando um mecanismo de rede (ou *sneakernet*) de compartilhamento de arquivos, para uma outra estação de trabalho para processamento adicional de conversão.

O digitalizador deve aceitar rolos de microfilme 35 mm, negativo ou positivo, com fotogramas em pé ou deitados, em vários graus de redução. Deve possuir um mecanismo de alimentação automático que permita operação não-assistida, além de ser equipado com sensores para a detecção do início e da borda de cada fotograma e para a correção automática da distorção dos fotogramas. Se possível, ele deve detectar o início e a borda de cada página nos casos em que se tenha duas páginas por fotograma. Em função do esforço dispensado na digitalização, o projeto de Yale procurará disponibilizar as imagens com a mais alta resolução tecnicamente possível. O digitalizador deve, contudo, ser capaz de escanear em diferentes níveis de resolução e, para prevenir estrangulamentos no trabalho, deve ser capaz de gerar as imagens em sua resolução máxima de saída a uma velocidade de dois segundos, ou menos, por fotograma. Finalmente, o digitalizador deve apresentar um conjunto aberto de interfaces que lhe permita trabalhar com estações de trabalho DOS, UNIX ou Macintosh. Um digitalizador para documentos em papel, de alta velocidade, com as mesmas características, deve estar à mão, para permitir comparações controladas de velocidade e qualidade nos casos em que a cópia em papel do material microfilmado ainda se encontra disponível.

Para trabalhar com as imagens digitalizadas, cada estação de trabalho, no processo de conversão, necessita da habilidade de processamento para ler e expandir a imagem comprimida de uma página e para comprimir e salvar, no disco, imagem de uma página exposta. Todas as estações de trabalho devem percorrer a seqüência de arquivos de imagem, para frente e para trás, e saltar para uma imagem específica a partir de uma outra qualquer. Cada estação de trabalho necessita, também, de um monitor de alta resolução capaz de exibir uma imagem de página legível. Cada estação de trabalho deveria proporcionar funções gerais de manipulação de imagem tais como *zoom*, redução, panorâmica, rolamento e rotação. Finalmente, todas as estações de trabalho no processo de conversão devem ter acesso a uma impressora laser local para a impressão de cópias de páginas selecionadas.

A estação de trabalho que controla o digitalizador e a qualidade de sua saída deve permitir ao operador tanto iniciar um processo de escaneamento totalmente automático, quanto interferir no mesmo, de forma clara e substancial. Ela deve possibilitar ao operador o ajuste e a determinação da resolução de escaneamento, produzindo, por exemplo, um conjunto de imagens a uma resolução elevada para impressão e um outro conjunto a uma resolução inferior para 'folheamento'. A estação de trabalho deve dar suporte às técnicas para realce automático de imagens, incluindo *thresholding* fixo e dinâmico e detecção de pico, que ajudam a melhorar o contraste entre texto e fundo, na impressão. Além disso, ela deve permitir ao operador cortar bordas de imagens e realizar outras funções de realce de imagem em tempo real, bem como reescanear completamente uma imagem quando necessário. Finalmente, a estação de trabalho deveria criar, automaticamente, arquivos TIFF a partir do digitalizador, comprimir as imagens escaneadas pela técnica bidimensional do grupo 4 do CCITT e atribuir nomes seqüenciais para os arquivos de imagens.



A aplicação de indexação no processo de conversão deve guiar o operador através de um processo que proporcione informação sobre certos aspectos chaves da estrutura e do conteúdo do conjunto de imagens que constituem o documento convertido. Em uma janela, deve exibir o conjunto de imagens e permitir ao operador percorrer, para frente e para trás, a seqüência de arquivos de imagens e ir diretamente até qualquer imagem específica. Em outra janela, a aplicação de indexação deve induzir o operador a verificar se a seqüência de imagens está correta e a fornecer um identificador único de documento sob o qual o conjunto de imagens possa ser armazenado como uma entidade única. Ao fornecer o identificador, a aplicação deveria checar na lista *online* dos identificadores existentes se não houve repetição.

Em uma terceira janela, a aplicação deve dar ao operador acesso ao catálogo local *online*, criar uma entrada apropriada para o documento em sua forma digital e indicar o identificador único através do qual o documento digital pode ser encontrado. Quando o trabalho bibliográfico estiver completo, o operador pode, opcionalmente, fornecer informações que relacionem a seqüência de imagens a uma notação de número de página e seqüência. O operador pode, também, indexar as divisões estruturais significativas dentro do documento e, então, poder salvar a estrutura e a informação de conteúdo em um arquivo de base de dados que, embora distinto, aponta para o conjunto de imagens que constituem o documento preservado. Se existem duas cópias de um documento em resoluções distintas, o arquivo de estrutura deve apontar precisamente para as cópias como exemplos distintos do mesmo documento com a mesma estrutura.

Uma vez que o arquivo de estrutura tenha sido criado, e até mesmo após ter sido permanentemente armazenado, o operador de indexação deve ser capaz de acessá-lo e modificá-lo. Além disso, uma vez que o operador tenha criado um índice remissivo de números de páginas para um documento específico, ele deve ser capaz, de imediato, de emitir um comando para ir até uma página específica. De forma similar, uma vez que o operador tenha criado um índice para a estrutura interna do documento, a aplicação deve permitir ao operador, com uma função especial de 'marcador de página', que assinalaria uma imagem específica e permitiria o retorno direto a ela a qualquer momento. A localização dos marcadores de página dos operadores, contudo, não deve ser salva na versão permanente do arquivo de estrutura.

Ao aplicar a função de confinamento (armazenagem permanente), o operador verifica se o processo de conversão está completo e se o conjunto de imagens digitais resultante deve então ser armazenado. A função assume, por motivos de segurança e de conveniência do usuário, que os documentos de imagem e que a base de dados de arquivos de estrutura a eles associada estejam armazenados em uma instalação de armazenagem central, acessível por rede e não por analogia ao microfilme, em armários localizados nas proximidades das estações de 'folheamento'. Nos casos em que o propósito da conversão não é criar uma cópia permanente do documento na forma digital e sim imprimi-lo, a função de armazenagem permanente pode encerrar o processo transferindo as imagens com o auxílio de meio removível ou através da rede, para a impressora digital de alta velocidade.

Com respeito à armazenagem em discos no processo de conversão, cada estação de trabalho que controla o digitalizador requer espaço de trabalho suficiente em disco magnético para que o operador possa armazenar e manipular as imagens escaneadas, não-comprimidas, de pelo menos dois livros (aproximadamente 600 megabytes). Em cada estação de trabalho que realiza as funções de indexação

e de armazenagem permanente, deve haver também um meio de armazenagem portátil, como um *drive* para disco óptico WORM (*write once, read many*) contendo o disco sobre o qual as imagens comprimidas do documento podem ser gravadas e, em seguida, fisicamente transportadas ao local de armazenagem permanente dos discos. Alternativamente, deve existir a possibilidade de gravar em uma instalação de armazenagem em massa através da rede de alta velocidade.

Armazenagem

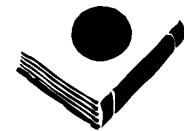
O propósito do subsistema de armazenagem é propiciar um meio (ou uma combinação de meios) de alta capacidade, confiável para armazenagem de uma coleção de documentos que foram preservados em forma digital ao subsistema de armazenagem que permita recuperar documentos tanto para impressão quanto para ‘folheamento’ em uma estação de trabalho computadorizada. Os componentes do subsistema de armazenagem incluem os dispositivos para armazenagem em massa, o controlador de armazenagem em massa, o servidor de base de dados de imagem e o arquivo de estrutura do documento.

Deve-se presumir que o processo de conversão crie cada arquivo de imagem a uma resolução de 400 dpi e que a forma comprimida de cada imagem ocupe 150 *kilobytes* na armazenagem e que cada livro convertido consista de 300 páginas. Por simples multiplicação, com estas suposições, concluímos que uma biblioteca digital para os 10 mil livros selecionados para conversão no projeto de Yale necessitará de aproximadamente 450 *gigabytes* para armazenagem em massa.

Considera-se que o disco óptico WORM é, atualmente, o meio mais eficiente, em termos de custo, para a armazenagem deste volume de informação¹⁹. Contudo, se uma segunda cópia de cada documento for criada a uma resolução inferior para fins de ‘folheamento’, as necessidades de armazenagem podem ser substancialmente menores para a segunda cópia (cerca de 125 GB) do que para a primeira e a frequência do uso pode ser suficiente para justificar os custos de armazenagem dos arquivos de imagem em disco magnético, que proporciona tempos de acesso mais rápidos que o meio óptico. Os discos ópticos WORM contendo as versões de maior resolução para impressão podem ser instalados em uma *jukebox*, ou mantidos fora de linha, o que exigiria um operador para intervir e instalá-lo quando solicitados. Os dados armazenados em meio magnético devem ser regularmente copiados em fita (*back up*), como precaução contra a perda de dados devido a falhas do disco; os dados armazenados pelos meios WORM devem ser periodicamente copiados para novos meios, na medida em que os formatos ou os equipamentos de acesso se tornem obsoletos e sejam substituídos.

O controlador de armazenagem em massa no subsistema proporciona uma interface entre a armazenagem em disco e o servidor de imagem. Ele traduz as solicitações de imagens em comandos que localizam e recuperam adequadamente os documentos de imagem armazenados no dispositivo de armazenagem magnético ou na *jukebox* para discos ópticos. Quando o documento estiver armazenado fora de linha, o controlador gerará uma solicitação ao operador para que instale o disco apropriado. Para solicitações de armazenagem de documentos de imagem, em vez de sua recuperação, o controlador localizará o espaço, armazenará os arquivos adequadamente e atualizará o diretório de arquivos.

¹⁹ Michael Lesk, op. cit., p. 306.



O arquivo de estrutura de documento contém aspectos de indexação, de informação, de estrutura e do conteúdo de cada documento de imagem. Ele é um componentes essencial da biblioteca de imagens e a informação nele contida deve acompanhar cada solicitação de documento; reciprocamente, qualquer novo documento de imagem que for armazenado deve conter um componente de índice para armazenagem no arquivo de estrutura. Esse arquivo é uma base de dados relacional em conformidade com a norma SQL. Sendo um arquivo *online*, ele requer cópia reserva (*backup*) regular.

O servidor de base de dados de imagem conecta o arquivo de estrutura do documento, através do controlador de armazenagem, aos arquivos de documentos de imagens. Ele é acessível pela rede e responde às solicitações de imagens, recuperando a informação relevante do arquivo de estrutura e os arquivos relevantes da biblioteca de imagens, preparando-os para entrega ao subsistema de acesso ou à impressora. O servidor também responde a solicitações para armazenagem de documentos de imagem assegurando que o arquivo de estrutura esteja atualizado e que os arquivos de imagem sejam transferidos ao dispositivo de armazenagem em massa apropriado. O servidor requer um volume substancial de memória armazenada (*cache*) para os documentos que entram e que saem. Uma função crucial do servidor neste processo é fornecer dados informativos de status a seus usuários, inclusive informação referente ao andamento da transmissão e erros na armazenagem ou na recuperação.

Acesso

O propósito do subsistema de acesso é dar ao usuário a possibilidade de recuperar e utilizar a biblioteca de documentos preservados na forma de imagem. Os componentes do subsistema incluem o servidor de apresentação de imagem e as estações de visualização de imagem. No caso mais simples do subsistema, documentos de imagem devem ser apresentados para visualização em um único tipo de estação de trabalho fornecido pela biblioteca e cuja seleção ela controla. O caso mais complicado ocorre quando a biblioteca provê acesso aos documentos de imagem para os pesquisadores de todo o campus, que utilizam diferentes tipos de estações de trabalho com capacidades de exibição de imagem variáveis. Mas, em qualquer caso, as funções mínimas proporcionadas por uma estação de visualização devem ser as mesmas. A estação deve permitir ao usuário recuperar um documento de imagem através de seu identificador único e exibir e ‘folhear’ o documento, dependendo da forma em que ele está descrito no arquivo de estrutura.

Para conseguir lidar com imagens, cada estação de trabalho requer a habilidade de processo para ler, expandir e exibir uma imagem de página comprimida. As estações requerem um monitor de alta resolução capaz de exibir uma imagem de página legível. Cada estação de trabalho deve proporcionar funções gerais de manipulação de imagem como *zoom*, redução panorâmica, rolamento e rotação. Finalmente, estações de visualização devem ser capazes de, se houver equipamento disponível, transferir imagens para impressoras a *laser* locais a impressão de cópias de páginas selecionadas.

A partir da estação de trabalho de acesso, o usuário deve ser capaz de consultar o catálogo *online* para determinar o identificador único de um documento. Determinado o identificador, o usuário deve, então, ser capaz de requisitar o documento ao servidor de imagem e tê-lo disponível para visualização na tela do monitor. Não havendo qualquer informação adicional de índice, o usuário deve ser capaz de percorrer o documento para frente e para trás, através da seqüência dos arquivos de imagem e de ir diretamente a qualquer imagem específica a partir de outra.

Contudo, se existir um índice de números de páginas, o usuário deve ser capaz de emitir um comando para ir a uma página específica. Da mesma forma, se houver um índice para o sumário, para as seções e os capítulos do documento, além de outras partes importantes de sua estrutura, o *software* deve possibilitar acesso imediato à imagem inicial de cada uma dessas divisões estruturais. A interface deve também proporcionar ao usuário uma função especial de marcador de página, que marque uma imagem específica e permita o retorno direto e imediato a essa imagem a qualquer momento; a localização dos marcadores de página dos usuários, contudo, existiria apenas durante a sessão de ‘folheamento’.

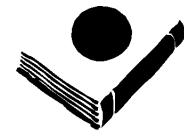
O servidor de apresentação de imagem separa essas diferentes funções de acesso, que todas as estações de visualização devem proporcionar em comum, das diferentes formas como os sistemas de operação das várias plataformas (DOS, UNIX, Macintosh) efetivamente as apresentam ao usuário. No início de uma sessão de ‘folheamento’, o servidor interrogará a estação de visualização específica para determinar seu tipo e suas capacidades. Pode ser que a estação não esteja adequadamente equipada para a visualização de imagens e, então, o servidor responderá neste sentido. Por outro lado, quando o servidor de imagem responder a uma solicitação por um documento, o servidor de apresentação será ativado. Ele usará o arquivo de estrutura do documento, que descreve aspectos críticos do documento, e, utilizando as conversões do Office Document Architecture, fornecerá um conjunto de alto nível de instruções sobre como o conteúdo do documento deve ser apresentado. O servidor traduz estas instruções para um formato e um conjunto de instruções de programas apropriados à estação de trabalho específica. Utilizando um *driver* obtido do servidor, a estação de visualização, por sua vez, processa as instruções e gera um formato de apresentação que permita ao usuário ‘folhear’ facilmente o documento de imagem.

Reprodução

O propósito do subsistema de reprodução é gerar cópias impressas de altíssima qualidade dos documentos selecionados, total ou parcialmente, que se encontram armazenados na forma de imagem digital. Os componentes do subsistema incluem a impressora, um servidor de impressão e um solicitador de impressão. A impressora deve ser capaz de aceitar como alimentação arquivos de imagem de resolução muito alta (600 pontos por polegada ou mais), devendo gerar cópias impressas com a mesma resolução. Para maior economia, ela deveria operar a velocidades bastante elevadas (100 páginas por minuto, ou mais, é um critério atual de referência) e apresentar capacidade interna de cotejamento e encadernação.

O servidor de impressão gerencia a fila de espera para a impressora, proporcionando vários controles de operação, incluindo a possibilidade de definir prioridades de execução, de informar sobre o andamento das mesmas e de cancelá-las. Ele deveria contar com uma capacidade de transferência de dados substancial e precisaria aceitar arquivos TIFF comprimidos de acordo com o padrão grupo 4 do CCITT. Se precisar reformatar tais arquivos para um formato interno, deve fazê-lo sem qualquer perda de qualidade ou informação. Para acomodar necessidades de impressão diretamente do subsistema de conversão, o servidor de impressão pode precisar de um drive WORM compatível para a entrada direta do documento a ser impresso.

Os usuários que queiram imprimir total ou parcialmente um documento específico o farão chamando um serviço de rede que submeta requisições de impressão. Um indivíduo pode solicitar uma cópia impressa de um documento de imagem através do identificador único especificado no catálogo



online. O serviço de requisição recuperará a informação de estrutura referente ao documento e pedirá ao mesmo para determinar quais as partes ou as páginas do documento a serem impressas e se é necessária a encadernação. O serviço de requisição de impressão deve possuir capacidade interna de contabilidade, de forma que os usuários possam ser cobrados pelas cópias que geraram. O uso geral do serviço de requisição em rede para cobrança presume que haja uma maneira de identificar os usuários e de verificar se eles são realmente quem dizem que são. Considerando a questão da identificação como tendo sido realizada, uma extensão natural do serviço de requisição de impressão será utilizada para coletar os encargos de *royalty* apropriados pela reprodução de documentos tais como artigos recentes de periódicos, que podem ser, eventualmente, adicionados à biblioteca digital.

Aspectos de exequibilidade

Como foi visto, um conjunto de normas técnicas está sendo elaborado para regular a arquitetura de sistemas de imagem digital. Além disso, vem tomando corpo um entendimento geral sobre a natureza e a inter-relação dos componentes do sistema de imagem digital para preservação em larga escala, em Yale, no qual o material convertido a partir do microfilme se torna, em formato de imagem, consideravelmente mais acessível e valioso à comunidade de pesquisadores.

O otimismo quanto à exequibilidade geral da criação de um sistema viável com base em uma biblioteca de imagens com 10 mil volumes não deve, contudo, obscurecer os riscos associados. Nem todos os aspectos dos componentes do sistema e de suas inter-relações são tão bem compreendidos quanto outras partes; há ainda muitas incertezas e aspectos desconhecidos sobre a tecnologia e sua operação. Invocando um princípio de projeto citado anteriormente — em que os dados criados e armazenados constituem o recurso principal de uma biblioteca em um sistema automatizado — e indagando de que maneira os dados estarão em maior risco nesta arquitetura específica, podemos identificar pelo menos quatro áreas de preocupação que devem limitar ou impedir o desenvolvimento do sistema e que, portanto, merecem especial atenção.

Em primeiro lugar, a área de maior preocupação é a integração dos dispositivos de armazenagem em massa. A tecnologia de armazenagem óptica WORM é, em geral, bem conhecida, mas formatos padronizados para o meio ainda não foram estabelecidos. Além disso, a experiência na indústria ainda não é muito rica no tocante à utilização atual da tecnologia e, em particular, à maneira de se otimizar a combinação de armazenagem magnética e óptica para aplicações específicas e, desta forma, melhorar o tempo de acesso e aliviar a disputa pelos *drives* em um ambiente multiusuário.

Diante das circunstâncias, a aplicação que a biblioteca espera desenvolver parece especial. As aplicações da tecnologia de reprodução digital de imagem estão ganhando espaço em ambientes de escritório, onde o documento modelo é uma pasta de arquivo que contém uns poucos itens e que alguém consulta rapidamente e segue em frente. O documento modelo na biblioteca, contudo, é o livro com algumas centenas de páginas, cujo texto alguém pode ler minuciosamente por algum tempo e, ocasionalmente, percorrer rapidamente para frente e para trás, para consultar pontos relacionados, citações, e assim por diante. Parece razoável prever que os diferentes padrões de uso de documentos na biblioteca sobrecarregarão os mecanismos de armazenagem em novas e diferentes maneiras, requerendo soluções que podem não estar, ainda, disponíveis. A questão é: a experiência corroborará esta hipótese ou não?

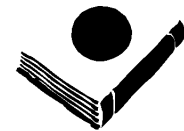
Em segundo lugar, a questão da renovação periódica dos dados em meios ópticos de armazenagem ainda causa preocupação. Centros de computação e fornecedores de computadores já atravessaram muitos ciclos de alterações técnicas no tocante aos meios magnéticos, exigindo a recópia periódica de dados. Os ciclos de mudança na tecnologia óptica, contudo, apenas começaram a atingir os grandes centros de produção de dados. Devido ao fato da tecnologia ser diferente e do volume de dados armazenado no disco óptico ser, em ordem de magnitude, muito superior ao volume armazenado nos meios magnéticos, os conhecidos paradigmas de renovação de dados podem não ser aplicados. Os fornecedores desempenharão um papel chave na difusão de soluções apropriadas e práticas; porém, quando interrogados, poucos são capazes de articular tais soluções.

A terceira área principal de preocupação é o denominado arquivo de estrutura do documento. À medida em que a biblioteca indexa seus documentos de imagem pelo número da página ou pela estrutura interna do documento, uma grande parte de seu trabalho e investimento vai sendo vinculada a este arquivo de estrutura. Os padrões ODA (*Office Document Architecture*) e SQL parecem apropriados para regular sua natureza e organização e para fazer com que os dados neles contidos sejam transferíveis para um outro sistema ou implementação quando necessário. Contudo, já está claro que este modo de investimento na biblioteca de imagens e nos vários índices incorporados no arquivo de estrutura de documento irá, sem dúvida, garantir a participação ativa da biblioteca de Yale não só na implementação específica da ODA, mas também em suas extensões e desenvolvimento.

Finalmente, um quarto aspecto de preocupação potencial emana do processo de controle de qualidade. A comunidade bibliotecária esperou durante anos que, dada a tecnologia correta, o microfilme fosse facilmente conversível à forma digital. O projeto de Yale gerará, certamente, bastante informação útil sobre a facilidade (ou dificuldade) técnica do processo de conversão. Porém, a qualidade do produto resultante do processo dependerá não somente dos mecanismos e da magia técnica do processo de conversão por si só, como também da qualidade do microfilme, da qualidade necessária às imagens digitais para servir como 'alimentação' para processos técnicos subsequentes, bem como do reconhecimento óptico de caracteres (OCR), e dos padrões de qualidade reivindicados pelos usuários que usarão as imagens digitais em seu trabalho de pesquisa. Será importante neste projeto, então, capturar e avaliar as formas pelas quais o acervo para preservação em microfilme possa ou deva ser alterado para gerar uma imagem digital melhor. Da mesma forma, será importante encontrar maneiras de avaliar a qualidade do processo de digitalização do ponto de vista dos processos técnicos afins, como o reconhecimento de caracteres, tornando-se absolutamente essencial avaliar o processo do ponto de vista da satisfação do usuário.

Plano de trabalho

O projeto de preservação digital de Yale será baseado em trabalhos relacionados, desenvolvidos na Biblioteca do Congresso norte-americano, na Biblioteca Nacional de Medicina, na Universidade de Cornell e em outras instituições bibliotecárias e arquivísticas, concentrando-se na conversão de microfilme para imagens digitais e almejando a construção de uma biblioteca digital com 10 mil volumes durante o processo. Grande parte do equipamento e do *software* de aplicação necessários à construção de um sistema em Yale já se encontra disponível no mercado, a partir dos fornecedores, ou encontra-se em estágio avançado de desenvolvimento pelos mesmos. A principal exceção é o gerenciador de apresentação de imagem descrito na seção anterior, para o qual a Yale conta com a mão-de-obra especializada de desenvolvimento e que será gerado como um produto específico do sistema de Yale.



Naturalmente, preocupações sobre vários aspectos ambíguos e incertos da arquitetura de sistemas de imagem digital acentuam os riscos do investimento em tecnologia relativamente nova e cara, na escala contemplada pelo projeto da Yale. Contudo, estas preocupações chamam, também, a atenção para as áreas de oportunidade que um projeto com o alcance e a extensão que tem o projeto da Yale oferece para o avanço de nossa compreensão acerca do potencial e dos limites da tecnologia, para o desenvolvimento de certos aspectos da tecnologia e, acima de tudo, para reduzir de modo mensurável os riscos do investimento para aquelas bibliotecas que venham a seguir o caminho da imagem digital. Para minorar os riscos do projeto e para ampliar suas possibilidades de gerar um retorno valioso, uma solução é um plano de implementação deliberado e cuidadoso com objetivos claros, resultados mensuráveis e uma abordagem em etapas, de forma que tudo não seja perdido ou esteja necessariamente em risco, caso surja algum problema ou obstáculo.

Objetivos do projeto

O projeto de preservação digital planejado para a Biblioteca da Universidade de Yale apresenta três objetivos maiores:

- identificar, desenvolver e avaliar os meios de criar e armazenar uma biblioteca digital de materiais preservados através da conversão de imagens armazenadas em microfilme para a forma digitalizada, numa escala de produção de 10 mil volumes;
- Identificar, desenvolver e avaliar os meios de propiciar, prover e melhorar a distribuição ou o acesso físico aos materiais preservados na forma de imagem digital, tanto de dentro quanto de fora da biblioteca;
- Identificar, desenvolver e avaliar os meios de prover, preservar e melhorar o acesso intelectual aos materiais da biblioteca na forma de imagem digital.

Escopo do projeto

O projeto será dividido em seis fases, ao longo de três anos (**Figura 7**). A primeira se estenderá do mês 1 ao 4 e estabelecerá a estrutura organizacional geral para o projeto. Durante a segunda fase, que se estenderá do mês 5 ao 8, será feita a instalação inicial do *hardware* e do *software*. A fase 3 estabelecerá um processo no nível de produção para a conversão de microfilme para imagens digitais. Ela terá início no mês 9 e terminará no mês 18. A quarta fase do projeto também terá início no mês 9, mas terminará no mês 21, quando os documentos e imagem digital se tornarem acessíveis aos usuários dentro da biblioteca de Yale. A quinta fase se estenderá do mês 22 ao 33 e estabelecerá o acesso aos documentos de imagem digital para usuários de fora da biblioteca de Yale. O projeto será concluído na sexta fase, durante os meses 34 ao 36, com a finalização da conversão dos 10 mil volumes, um ajuste final e a avaliação de todos os componentes do sistema.

Em cada estágio, um relatório será gerado para assumir os resultados até então alcançados e para avaliar o mérito geral do projeto. Devido ao fato de que a tecnologia de imagem digital que está sendo investigada e demonstrada ainda está sujeita a rápidas alterações e desenvolvimento, é possível, embora pouco provável, que os resultados daqui vislumbrados emergem em algum outro lugar ou que as capacidades da própria tecnologia ultrapassem as pretensões em que os projetos se fundamenta. Desta forma, cada relatório tratará o final daquela fase como um marco para perguntar e responder a seguinte questão: é válido continuar este projeto?

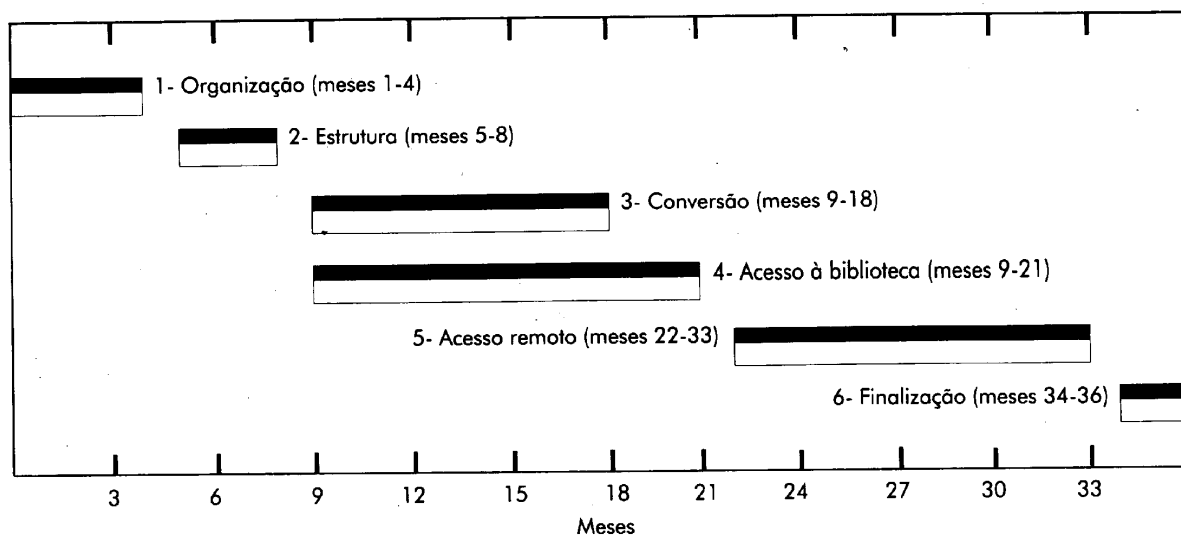


Figura 7. Fases do projeto

Fase 1 - Organização

Objetivo - O objetivo da fase 1 é criar a estrutura organizacional geral para o projeto e preparar o trabalho para as fases subsequentes.

Cronograma - Meses 1-4

Atividades - Durante a fase 1, um Comitê Diretor será criado para prover a supervisão geral e a direção para o projeto. Ele será constituído, ao menos inicialmente, pelo diretor da Biblioteca da Universidade, pelo diretor substituto, pelos bibliotecários ligados à Biblioteca para serviços técnicos, atendimento ao público e desenvolvimento de coleções — pelo chefe do Departamento de Preservação, pelo diretor de Computação e Sistemas de Informação da Universidade, pelo diretor acadêmico de Computação, pelo gerente do projeto e por quatro membros selecionados da equipe. O Comitê pode ajustar sua composição ao longo do tempo, se necessário.

A agenda inicial do Comitê incluirá os seguintes itens:

- estabelecer os critérios para identificar que porções da coleção de microfilme para preservação serão convertidas em imagens digitais;
- auxiliar o gerente do projeto e membros da equipe do projeto, no estabelecimento de relações de parceria adequadas, com os fornecedores de equipamento e aplicações para o processamento digital de documentos;
- rever o orçamento proposto para o projeto e auxiliar na identificação e obtenção de fontes de financiamento.

O gerente do projeto será o chefe do escritório de sistemas da biblioteca. A equipe do projeto se constituirá, inicialmente, pelo chefe do Departamento de Preservação da Biblioteca e pelo diretor associado de Sistemas, Tecnologias e Planejamento do Departamento de Computação e Sistemas de Informação. Durante esta fase, a equipe do projeto:



- identificará o *hardware* e o *software* específicos que serão adquiridos para o projeto;
- estabelecerá relações de parceria adequadas com os fornecedores de equipamento e aplicações para o processamento digital de documentos que serão adquiridos no projeto;
- preparará um plano da composição do quadro de profissionais envolvidos para todas as fases subsequentes do projeto;
- preparará um orçamento para o projeto inteiro e trabalhará com o Comitê Diretor e outras fontes apropriadas para a obtenção de financiamentos.

Pré-requisitos - Este documento, *Do microfilme à imagem digital*, proporcionará a base e os fundamentos de planejamento para o trabalho inicial do Comitê Diretor e da equipe do projeto.

Produtos

- um orçamento para o projeto;
- um plano de custeio;
- um relatório resumindo os resultados do trabalho nesta fase.

Fase 2 - Estrutura

Objetivo - O objetivo da fase 2 é iniciar a instalação do ambiente de trabalho necessário ao projeto.

Cronograma - Meses 5-8.

Atividades - A equipe do projeto:

- aplicará os critérios estabelecidos pelo Comitê Diretor e iniciará a seleção dos segmentos da coleção de microfilme a serem convertidos;
- criará e iniciará a implementação de um plano para avaliar se os critérios utilizados na seleção de material para conversão criarão uma biblioteca digital que os usuários da Biblioteca considerem intelectualmente valiosa;
- contratará os profissionais necessários;
- adquirirá, instalará e testará o *hardware* e o *software* para o subsistema de conversão;
- adquirirá, instalará e testará a impressora digital de alta velocidade e o servidor de impressão.
- adquirirá, instalará e testará o componente do arquivo de estrutura de documento do subsistema de armazenagem e começará a adquirir, instalar e testar os outros componentes do subsistema de armazenagem;
- preparará o projeto e as especificações iniciais para o servidor de apresentação de imagem.

Além disso, a equipe firmará um contrato com uma agência de prestação de serviços para converter, de microfilme para a forma de imagem digital, até 10 volumes selecionados aleatoriamente. O contrato estabelecerá a linha base de custos para comparação com os custos de processamento na própria instituição. Fornecerá, também, experiência na formulação e aplicação de critérios de aceitação para o processo de controle de qualidade dentro da instituição. Presumivelmente, neste estágio, os documentos serão desenvolvidos na forma de imagem, sem qualquer indexação.

Pré-requisitos - Para que esta fase se inicie, deve haver um orçamento aprovado, provido dos fundos necessários para cobrir os custos de pessoal, da aquisição de *hardware* e *software* e da agência de prestação de serviços.

Produtos

- dez volumes estarão disponíveis na forma digital e prontos para o processo de indexação dentro do subsistema de conversão;
- um relatório resumindo os resultados do trabalho nesta fase, incluindo uma análise da experiência com a agência de prestação de serviços e um plano para avaliar sistematicamente, durante toda a duração o projeto, os méritos da conversão dentro da própria instituição *versus* a utilização da agência de prestação de serviços.

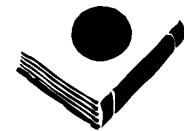
Fase 3 - Conversão

Objetivo - O objetivo da fase 3 é estabelecer um processo em nível de produção para a conversão de microfilme para imagens digitais.

Cronograma - Meses 9-18.

Atividades - As principais tarefas da equipe do projeto nesta fase incluem:

- desenvolver procedimentos e treinar o(s) operador(es) no uso do digitalizador e da estação de trabalho de controle de qualidade, incluindo suas características de realce de imagens;
- selecionar para a digitalização, no início desta fase, uma amostra de carretéis de microfilme que reflita as variações na qualidade do processo de microfilmagem. Planejar um método para utilizar esta amostra tanto para praticar o processo de controle de qualidade quanto para iniciar a formulação de conclusões sobre as maneiras como as técnicas de microfilmagem podem afetar a qualidade do processo de digitalização;
- planejar e implementar um padrão de controle de qualidade que considere, em especial, a possibilidade futura de aplicação da tecnologia de reconhecimento de caracteres às imagens geradas neste processo de conversão;
- criar um processo contínuo para a avaliação da satisfação do usuário com os processos e as normas técnicas de controle de qualidade;
- rever o projeto do índice de número de página e do índice de estrutura do documento para assegurar que conexões podem ser criadas, no futuro, entre estes índices e as notas de conteúdo suplementares e que possam estar armazenadas em algum outro local, em formato legível por computador;
- desenvolver procedimentos e treinar o(s) operador (es) no uso de cada uma das quatro diferentes funções de indexação de documentos. O(s) operador(es) no uso de cada uma das quatro diferentes funções de indexação de documentos. O(s) operador(es) deveria(m) ser capaz(es) de: 1) atribuir um identificador único ao conjunto de imagens convertidas que constituiriam uma entidade documental única; 2) indexar o documento imagem por número de página; 3) indexar o documento imagem pela estrutura interna do documento; 4) criar uma entidade bibliográfica no catálogo *online* da Yale que descreva o documento imagem e o identifique na biblioteca de imagem digital através de seu identificador único;



- planejar métodos de estruturação e alteração do fluxo de trabalho de indexação para isolar os custos de cada uma das quatro formas de indexação de documentos;
- desenvolver procedimentos e treinar o(s) operador(es) para confinar o documento imagem e os índices do documento à armazenagem em arquivo permanente utilizando, neste estágio, um *drive* portátil WORM;
- revisar os critérios para a seleção de materiais de preservação de alta utilização dos quais uma cópia impressa será encadernada e devolvida à estante;
- desenvolver procedimentos e treinar o(s) operador(es) para transferir as imagens de documentos selecionados diretamente para a impressora de documentos de imagem de alta qualidade. O modo de transmissão preferível é através da rede de alta velocidade;
- planejar um método de estruturação do fluxo de trabalho para isolar os custos de digitalização, controle de qualidade e impressão, sem qualquer indexação;
- planejar um método e um procedimento contínuo para comparar os custos de digitalização, controle de qualidade e indexação na própria instituição com os custos correspondentes, a nível de agência de prestação de serviços;
- indexar e confinar à armazenagens os 10 volumes digitalizados pela agência de prestação de serviços na fase 2;
- digitalizar, realçar, indexar e confinar à armazenagem, em arquivo permanente, 490 volumes.

Pré-requisitos - Para que esta fase possa começar, o subsistema de conversão, a impressora, o servidor de impressão e o componentes de arquivo de estrutura de documento do subsistema de armazenagem devem todos estar instalados e testados.

Aspectos a serem investigados - O trabalho nesta fase é concebido para produzir medição criteriosa de qualidade e custo ao longo de várias dimensões, incluindo:

- os possíveis efeitos da qualidade do microfilme sobre a qualidade da digitalização;
- as implicações de um possível futuro reconhecimento de caracteres no processo de controle de qualidade na digitalização do microfilme;
- a satisfação do usuário com os padrões e processos de controle de qualidade;
- os custos relativos de digitalização, controle de qualidade e indexação na própria instituição e em uma agência de prestação de serviços;
- os custos variáveis dos diferentes níveis de indexação;
- os custos da digitalização e impressão simples, sem indexação.

Será importante coletar e monitorar estas várias medidas em uma base contínua durante todo o projeto.

Produtos

- operação do subsistema de conversão em níveis de produção;
- 500 volumes digitalizados e completamente indexados;

- um relatório resumindo os resultados do trabalho nesta fase, incluindo uma análise dos custos.

Fase 4 - Acesso à Biblioteca

Objetivo - O objetivo da fase 4 é tornar a crescente biblioteca de imagem digital acessível aos usuários através de estações de trabalho localizadas dentro da biblioteca de Yale.

Cronograma - Meses 9-21.

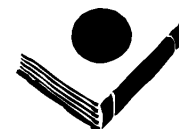
Atividades - As principais tarefas da equipe do projeto nesta fase incluem:

- completar a aquisição, instalação e teste do subsistema de armazenagem;
- especificar, adquirir, instalar e testar as estações de acesso na biblioteca;
- desenvolver e testar um protótipo do servidor de apresentação de imagem capaz de interagir com as estações de acesso, na biblioteca;
- projetar e implementar um estudo para medir o valor, para os usuários da biblioteca, de cada uma das ferramentas de índice disponíveis para o documento digital, em comparação com o acesso ao microfilme;
- criar e implementar um plano para fazer cópia reserva (*back up*) dos dados armazenados sobre meios magnéticos e para, periodicamente, renovar os dados armazenados nos meios ópticos e magnéticos;
- transferir todos os volumes digitalizados para o sistema de armazenagem permanente;
- conceber um método para estimar os custos da adição do subsistema de armazenagem e de manutenção a uma base contínua;
- estabelecer e implementar métodos e procedimentos contínuos para estimar os efeitos da carga de acesso sobre a performance do subsistema de armazenagem de imagem;
- conceber e implementar métodos e procedimentos contínuos para monitoramento e ajuste dos defeitos da transmissão de imagens sobre a performance da rede;
- conceber métodos para estimar os custos da criação e manutenção do subsistema de acesso;
- adquirir, instalar e testar o solicitador de impressão;
- conceber métodos para estimar os custos da criação e manutenção do subsistema de impressão;
- após o término da fase 2 e durante os meses 16-18, digitalizar, realçar, indexar e confinar à armazenagem permanente 1.300 volumes.

Pré-requisitos - Para que esta fase se inicie, a impressora, o servidor de impressão e o componente de arquivo de estrutura de documento do subsistema de armazenagem devem estar instalados e testados. O projeto inicial e as especificações para o servidor de apresentação de imagem devem também estar completos.

Aspectos a ser investigados - O trabalho nesta fase é concebido para produzir medição criteriosa de qualidade e custo ao longo de várias dimensões, incluindo:

- os custos da criação e manutenção do subsistema de armazenagem;



- os custos de confecção de cópia reserva (*backup*) dos dados armazenados sobre disco magnético e da renovação periódica dos dados armazenados em meios óptico e magnético;
- os custos da criação e manutenção do subsistema de acesso;
- os custos da criação e manutenção do subsistema de impressão;
- os efeitos da carga de acesso sobre o subsistema de armazenagem;
- os efeitos da carga de transmissão sobre a rede;
- o valor dos índices de documento para os usuários na biblioteca;

Será importante coletar e monitorar estas várias medidas em uma base contínua durante todo o projeto.

Produtos

- Subsistema de armazenagem de imagens operando em níveis de produção;
- Subsistema de impressão operando em níveis de produção;
- Protótipo do gerenciador de apresentação de imagem;
- Subsistema de acesso operando em níveis de produção na biblioteca;
- Servidor de solicitação de impressão;
- 1.300 volumes digitalizados e completamente indexados;
- Um relatório resumindo os resultados do trabalho nesta fase, incluindo uma análise de custos.

Fase 5 - Acesso remoto

Objetivo - O objetivo da fase 5 é estabelecer o acesso aos documentos de imagem digital de fora da biblioteca de Yale.

Cronograma - Meses 22-23

Atividades - As principais tarefas da equipe do projeto nesta fase incluem:

- estender a operação do servidor de apresentação de imagens, de forma a torná-lo capaz de interagir com tipos selecionados de estações de trabalho fora da biblioteca;
- tornar o subsistema de acesso às imagens disponível, através do servidor de apresentação de imagens, a tipos selecionados de estações de trabalho conectadas à rede do campus;
- conceber métodos para estimar os custos iniciais e os custos contínuos da introdução do acesso ao sistema de imagem de fora da biblioteca;
- planejar e implementar um estudo para medir o valor, para os usuários de fora da biblioteca, de cada uma das ferramentas de índice disponíveis para o documento digital;
- digitalizar, realçar, indexar e confinar à armazenagem permanente 7.800 volumes.

Pré-requisitos - Para que esta fase se inicie, o protótipo do servidor de apresentação de imagem deve estar completo e operacional.

Aspectos a ser investigados - O trabalho nesta fase é projetado para produzir medições criteriosas do valor dos índices de documento para usuários de fora da biblioteca. Será importante coletar e monitorar estas medidas em uma base contínua durante esta e a última fase do projeto.

Produtos

- gerenciador de apresentação de imagem em nível de produção;
- operação do subsistema de acesso a partir de estações de trabalho fora da biblioteca;
- 7.800 volumes digitalizados e completamente indexados;
- um relatório resumindo os resultados do trabalho nesta fase, incluindo uma análise de custos.

Fase 6 - Finalização

Objetivo - O objetivo da fase 6 é completar a conversão dos 10 mil volumes e realizar um ajuste e uma avaliação finais de todos os componentes do sistema.

Cronograma - Meses 34-36.

Atividades - As tarefas principais da equipe do projeto nesta fase incluem:

- rever e avaliar todas as medições de qualidade e performance em cada um dos subsistemas.
- digitalizar, realçar, indexar e confinar à armazenagem permanente 400 volumes.

Pré-requisitos - Para que esta fase se inicie, a versão de produção do servidor de apresentação de imagens deve estar completa e operacional.

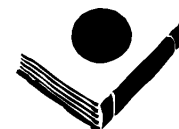
Produtos

- um relatório final;
- 400 volumes digitalizados e completamente indexados.

Conclusão

A biblioteca da Universidade de Yale prevê um futuro em que a tecnologia de imagem digital constituirá uma ferramenta fundamental no processo de preservação dos materiais em deterioração de suas ricas e valiosas coleções. Justamente no momento em que a deterioração de documentos compostos de papel ácido está se acelerando a uma taxa alarmante em Yale e em outras grandes bibliotecas de pesquisa por todo os Estados Unidos, a arquitetura de sistemas de imagem digital está se tornando mais conhecida, e os componentes do sistema estão cada vez mais bem integrados e prontamente disponíveis. A Biblioteca do Congresso norte-americano, A Biblioteca Nacional de Medicina, a Universidade Cornell e outras instituições bibliotecárias e arquivísticas têm fornecido, de diferentes formas, grandes contribuições ao desenvolvimento da tecnologia de imagem digital e à sua aplicação para preservar e melhorar o acesso geral ao manancial nacional e internacional de conhecimento registrado. Este é o momento certo para outro grande passo no sentido de fazer da tecnologia de imagem digital uma ferramenta prática para bibliotecas.

O projeto de preservação digital de Yale será fundamentado em trabalho de igual natureza aos desenvolvidos em outras instituições, particularmente na Universidade de Cornell. Concentrando-se



na conversão de microfilme para imagens digitais e almejando a construção de uma biblioteca digital com 10 mil volumes. Grande parte do equipamento e do *software* de aplicação necessários à construção de um sistema em Yale já está disponível no mercado, ou em estágio avançado de desenvolvimento pelos fornecedores. Contudo, o sistema de processamento de imagens digitais para preservação da Yale implicará em um grau elevado de capital e, particularmente na fase de indexação do processo de conversão, de trabalho. Seu sucesso exigirá auxílio considerável de agências de financiamento e parcerias significativas e produtivas com membros da comunidade de fornecedores. O resultado para todos os participantes do projeto será um corpo substancial de informações sobre os custos e benefícios da tecnologia de imagem digital. As bibliotecas, e os fornecedores que com elas trabalham, dependerão de tais informações, no futuro, para tomar decisões prudentes e práticas sobre a incorporação da tecnologia aos materiais nela contidos.

Notas

Agradecimentos: Na preparação deste relatório de planejamento, eu recebi muita ajuda de várias pessoas em várias instituições privadas e universitárias. Elas são muito numerosas para serem individualmente citadas; eu sou muito grato a todas elas. Contudo, quero agradecer a meus colegas do Escritório de Sistemas da Biblioteca da Universidade de Yale, Merri Beth Lavagnino e Greg Kaisen, por sua considerável ajuda. Agradeço, também, à Commission of an Access Preservation por seu apoio através da contratação da biblioteca da Universidade de Yale para elaborar este relatório.

Donald J. Waters
Chefe do Escritório de Sistemas
Biblioteca da Universidade de Yale
New Haven, Connecticut 06520

O Projeto Conservação Preventiva em Bibliotecas e Arquivos - CPBA

está sediado no
Arquivo Nacional
Rua Azeredo Coutinho 77, sala 605 - C
CEP 20230-170 Rio de Janeiro - RJ
Tel/Fax: (21) 2253-2033
www.cpba.net
www.arquivonacional.gov.br

The Council on Library and Information Resources - CLIR
(incorporando a antiga ***Commission on Preservation and Access***)

1755 Massachusetts Avenue, NW, Suite 500
Washington, DC 20036
Tel: (202) 939-4750
Fax: (202) 939-4765
www.clir.org

Títulos Publicados

Armazenagem e manuseio

1. Métodos de armazenamento e práticas de manuseio
2. A limpeza de livros e de prateleiras
3. A escolha de invólucros de qualidade arquivística para armazenagem de livros e documentos
4. Invólucros de cartão para pequenos livros
5. A jaqueta de poliéster para livros
6. Suporte para livros: descrição e usos
7. Montagens e molduras para trabalhos artísticos e artefatos em papel
8. Mobiliário de armazenagem: um breve resumo das opções atuais
9. Soluções para armazenagem de artefatos de grandes dimensões

Conservação

10. Planificação do papel por meio de umidificação
11. Como fazer o seu próprio passe-partout
12. Preservação de livros de recortes e álbuns
13. Manual de pequenos reparos em livros

Melo Ambiente

14. Temperatura, umidade relativa do ar, luz e qualidade do ar: diretrizes básicas de preservação
15. A proteção contra danos provocados pela luz
16. Monitoramento da temperatura e umidade relativa
17. A proteção de livros e papéis durante exposições
18. Isopermas: uma ferramenta para o gerenciamento ambiental
19. Novas ferramentas para preservação-avaliando os efeitos ambientais a longo prazo sobre coleções de bibliotecas e arquivos

Emergências

20. Planejamento para casos de emergência
21. Segurança contra as perdas: danos provocados por água e fogo, agentes biológicos, roubo e vandalismo
22. Secagem de livros e documentos molhados
23. A proteção de coleções durante obras
24. Salvamento de fotografias em casos de emergência
25. Planilha para o delineamento de planos de emergência
26. Controle integrado de pragas
27. A proteção de livros e papel contra o mofo
28. Como lidar com uma invasão de mofo: instruções em resposta a uma situação de emergência
29. Controle de insetos por meio de gases inertes em arquivos e bibliotecas

Planejamento

30. Planejamento para preservação
31. Políticas de desenvolvimento de coleção e preservação
32. Planejamento de um programa eficaz de manutenção de acervos
33. Desenvolvimento, gerenciamento e preservação de coleções
34. Seleção para preservação: uma abordagem materialística
35. Considerações complementares sobre: "Seleção para preservação: uma abordagem materialística"
36. Implementando um programa de reparo e tratamento de livros
37. Programa de Planejamento de Preservação: um manual para auto-instrução de bibliotecas

Edifício/Preservação

38. Considerações sobre preservação na construção e reforma de bibliotecas: planejamento para preservação

Fotografias e filmes

39. Preservação de fotografias: métodos básicos para salvaguardar suas coleções
40. Guia do Image Permanence Institute (IPI) para armazenamento de filmes de acetato
41. Indicações para o cuidado e a identificação da base de filmes fotográficos

Registros sonoros e fitas magnéticas

42. Armazenamento e manuseio de fitas magnéticas
43. Guarda e manuseio de materiais de registro sonoro

Reformatação

44. O básico sobre o processo de digitalizar imagens
45. Microfilme de preservação: plataforma para sistemas digitais de acesso
46. O processo decisório em preservação e fotocopiagem para arquivamento
47. Controle de qualidade em cópias eletrostáticas para arquivamento
48. Microfilmagem de preservação: um guia para bibliotecários e arquivistas
49. Do microfilme à imagem digital
50. Uma abordagem de sistemas híbridos para a preservação de materiais impressos
51. Requisitos de resolução digital para textos: métodos para o estabelecimento de critérios de qualidade de imagem
52. Preservação no universo digital
53. Manual do RLG para microfilmagem de arquivos