

CONARQ

**RECOMENDAÇÕES PARA DIGITALIZAÇÃO DE DOCUMENTOS
ARQUIVÍSTICOS PERMANENTES**

**abril
2010**

Créditos

Grupo Técnico de Redação do Arquivo Nacional

Adriana Cox Hollós	Conselho Nacional de Arquivos - CONARQ
Alex Pereira de Holanda	Coordenação de Preservação de Acervo - COPAC
Ana Celeste Indolfo	Coordenação Geral de Gestão de Documentos - COGED
Carlos Augusto Silva Ditadi	Coordenação Geral de Gestão de Documentos - COGED
Mauro Domingues de Sá	Coordenação de Preservação do Acervo - COPAC

Grupo de Trabalho do Arquivo Nacional

Adriana Cox Hollós	Conselho Nacional de Arquivos - CONARQ
Alex Pereira de Holanda	Coordenação de Preservação de Acervo - COPAC
Ana Celeste Indolfo	Coordenação Geral de Gestão de Documentos - COGED
Carlos Augusto Silva Ditadi	Coordenação Geral de Gestão de Documentos - COGED
Beatriz Moreira Monteiro	Coordenação de Documentos Escritos - CODES
Gerson Pereira	Coordenação de Preservação do Acervo - COPAC
José Luiz Macedo de Farias Santos	Coordenação de Documentos Audiovisuais e Cartográficos - CODAC
Leonardo Vicente de Pontes	Coordenação de Consulta ao Acervo - COCAC
Mauro Domingues de Sá	Coordenação de Preservação do Acervo - COPAC
Mauro Lerner Markowski	Coordenação de Documentos Escritos - CODES
Sérgio Miranda de Lima	Coordenação de Documentos Audiovisuais e Cartográficos - CODAC
Valéria Maria Morse Alves	Coordenação de Consulta ao Acervo - COCAC

Revisão de Referências Técnicas

Elizabeth da Silva Maçulo	Coordenação Geral de Gestão de Documentos - COGED
---------------------------	---

**CONSELHO NACIONAL DE ARQUIVOS
CONARQ**

**RECOMENDAÇÕES PARA DIGITALIZAÇÃO DE DOCUMENTOS
ARQUIVÍSTICOS PERMANENTES**

**Abril
2010**

SUMÁRIO

- 1. APRESENTAÇÃO**
- 2. PÚBLICO-ALVO**
- 3. O QUE ESSA RECOMENDAÇÃO ABORDA**
- 4. O QUE ESSA RECOMENDAÇÃO NÃO ABORDA**
- 5. O QUE É A DIGITALIZAÇÃO**
- 6. POR QUE DIGITALIZAR**
- 7. PROJETO DE DIGITALIZAÇÃO**
- 8. CAPTURA DIGITAL DA IMAGEM**
- 9. ACESSO**
- 10. ASPECTOS GERAIS PARA SEGURANÇA, ARMAZENAMENTO E PRESERVAÇÃO DOS REPRESENTANTES DIGITAIS**
- 11. UTILIZAÇÃO DE TERCEIRIZAÇÃO DE SERVIÇOS DE DIGITALIZAÇÃO**
 - ANEXO I - MODELO DE VERIFICAÇÃO DA QUALIDADE DO REPRESENTANTE DIGITAL**
 - ANEXO II - REFERENCIAS**

1. APRESENTAÇÃO

Essa recomendação visa auxiliar as instituições detentoras de acervos arquivísticos de valor permanente¹, na concepção e execução de projetos e programas de digitalização.

A digitalização de acervos é uma das ferramentas essenciais ao acesso e à difusão dos acervos arquivísticos, além de contribuir para a sua preservação, uma vez que restringe o manuseio aos originais, constituindo-se como instrumento capaz de dar acesso simultâneo local ou remoto aos seus representantes digitais² como os documentos textuais, cartográficos e iconográficos em suportes convencionais, objeto desta recomendação.

A adoção de um processo de digitalização implica no conhecimento não só dos princípios da arquivologia, mas também no cumprimento das atividades inerentes ao processo, quais sejam a captura digital, o armazenamento e a disseminação dos representantes digitais. Isto quer dizer que os gestores das instituições arquivísticas e os demais profissionais envolvidos deverão levar em consideração os custos de implantação do projeto de digitalização, compreendendo que um processo como este exige necessariamente um planejamento com previsão orçamentária e financeira capazes de garantir a aquisição, atualização e manutenção de versões de *software* e *hardware*, a adoção de formatos de arquivo digitais e de requisitos técnicos mínimos que garantam a preservação e a acessibilidade a curto, médio e longo prazos dos representantes digitais gerados.

Devido à natureza complexa de um ambiente tecnológico de rápidas mudanças e, em geral, de custos elevados para sua implementação e manutenção, sugere-se elaborar projetos cooperativos com outras organizações possuidoras de infra-estrutura tecnológica e pessoal especializado, que ofereçam estrutura tecnológica adequada à captura digital, ao acesso e ao armazenamento com garantias de segurança e preservação a longo prazo.

Importante destacar que as ações de digitalização não devem ser realizadas em detrimento das ações de conservação convencional dos acervos custodiados por instituições arquivísticas, por serem inalienáveis e imprescritíveis, conforme preconiza o artigo 10º da Lei Federal nº 8.159/1991.

Os requisitos contidos nesta recomendação visam garantir que um projeto de digitalização de documentos arquivísticos de valor permanente contemple a geração de representantes digitais com qualidade arquivística, fidelidade ao documento original e capacidade de interoperabilidade, evitando-se ao longo do tempo a necessidade de se refazer a digitalização³, além de garantir a satisfação das necessidades de uso dos usuários finais.

1 Cf. art. 7º § 3º e art. 10º da Lei nº 8.159, de 08 de janeiro de 1991. Disponível em: <<http://www.conarq.arquivonacional.gov.br>>

2 **Representante digital** - (*digital surrogate*) - Nos termos dessa Recomendação é a representação em formato de arquivo digital de um documento originalmente não digital. É uma forma de diferenciá-lo do documento de arquivo nascido originalmente em formato de arquivo digital (*born digital*).

3 Para maiores informações sobre critérios de concepção de projetos de digitalização de documentos arquivísticos, recomendamos os Cadernos Técnicos do Projeto de Conservação Preventiva de Bibliotecas e Arquivos - CBPA: nº 44 - **O básico sobre o processo de digitalizar imagens**; nº 45 - **Microfilme de preservação: plataforma para sistemas digitais de acesso**; nº 46 - **O processo decisório em preservação e fotocopiagem para arquivamento**; nº 49 - **Do microfilme à imagem digital**, e nº 50 - **Uma abordagem de sistemas híbridos para a preservação de materiais impressos**. Disponíveis em: <<http://siarq49.siarq.unicamp.br/cpba>>.

2. PÚBLICO ALVO

As instituições arquivísticas, integrantes do Sistema Nacional de Arquivos - SINAR, e demais responsáveis pela custódia de acervos documentais, públicos ou privados, considerados de valor permanente. Outras organizações podem utilizar essa Recomendação como referência.

3. O QUE ESSA RECOMENDAÇÃO TRATA

- Captura digital⁴ em imagem de documentos planos e encadernados: impressos, manuscritos, mapas, plantas, desenhos, gravuras, cartazes, microformas, diapositivos, negativos, cópias e ampliações fotográficas;
- Padrões e boas práticas mínimas para a captura digital de imagens; produção de matrizes e derivadas, identificação do representante digital e controle de qualidade;
- Formatos digitais para representantes digitais matrizes e derivados;
- Metadados técnicos;
- Boas práticas gerais para armazenamento, segurança e preservação dos representantes digitais;
- Utilização de serviços terceirizados para a captura digital, armazenamento e acesso aos representantes digitais.

4. O QUE ESSA RECOMENDAÇÃO NÃO TRATA

- Seleção do acervo a ser digitalizado;
- Descrição arquivística e política de acesso e uso dos documentos arquivísticos;
- Digitalização de acervo sonoro e de imagens em movimento;
- Preservação em longo prazo daqueles documentos originalmente criados e mantidos em ambiente eletrônico (*born digital*);⁵
- Capacitação e qualificação de pessoal;
- Aspectos legais relativos aos direitos autorais, de acesso, privacidade e sigilo.

5. O QUE É A DIGITALIZAÇÃO

Entendemos a digitalização como um processo de conversão dos documentos arquivísticos em formato digital, que consiste em unidades de dados binários, denominadas de *bits* - que são 0 (zero) e 1 (um), agrupadas em conjuntos de 8 *bits* (*binary digit*) formando um

4 **Captura digital.** Nos termos dessa recomendação, significa a conversão em imagem, por dispositivo eletrônico (escâner), para o formato digital de um documento originalmente não digital. Ver nota 2, Representante Digital.

5 Indicamos algumas boas práticas em: **Aspectos Gerais para Segurança, Armazenamento e Preservação dos Representantes Digitais.**

byte, e com os quais os computadores criam, recebem, processam, transmitem e armazenam dados.

De acordo com a natureza do documento arquivístico original, diversos dispositivos tecnológicos (*hardware*) e programas de computadores (*software*) serão utilizados para converter em dados binários o documento original para diferentes formatos digitais. No entanto, o produto dessa conversão não será igual ao original e não substitui o original que deve ser preservado.

A digitalização, portanto é dirigida ao acesso, difusão e preservação do acervo documental.

6. POR QUE DIGITALIZAR

- Contribuir para o amplo acesso e disseminação dos documentos arquivísticos por meio da Tecnologia da Informação e Comunicação;
- Permitir o intercâmbio de acervos documentais e de seus instrumentos de pesquisa por meio de redes informatizadas;
- Promover a difusão e reprodução dos acervos arquivísticos não digitais, em formatos e apresentações diferenciados do formato original;
- Incrementar a preservação e segurança dos documentos arquivísticos originais que estão em outros suportes não digitais, por restringir seu manuseio.

7. PROJETO DE DIGITALIZAÇÃO

Antes do início do projeto, pressupõe-se que o tratamento arquivístico dos documentos e a avaliação e seleção dos conjuntos documentais a serem digitalizados, já tenham sido desenvolvidos, e que o acervo arquivístico selecionado tenha sido previamente higienizado, identificado e organizado (arranjo, descrição e indexação) .

A fim de se gerar um representante digital fiel ao documento original, deve-se identificar o menor caractere (linha, traço, ponto, mancha de impressão) a ser digitalizado para a determinação da resolução óptica que garantirá sua legibilidade na versão digital. Para a definição do tipo de equipamento de captura digital a ser utilizado, deve-se observar os tipos documentais existentes no acervo e sua quantificação, além das características físico-químicas de cada tipo de documento, para reduzir os riscos⁶ à integridade física do original.

O processo de digitalização deverá ser realizado, preferencialmente, nas instalações das instituições detentoras do acervo documental, evitando seu transporte e manuseio inadequados, e a possibilidade de danos causados por questões ambientais, roubo ou extravio. Na eventualidade de terceirização ou contratação de serviços ver: **Utilização de serviços terceirizados de digitalização e correlatos.**

Recomenda-se a digitalização de conjuntos documentais integrais, como fundos/coleções ou séries. No entanto, é possível digitalizar itens documentais isolados,

6 Recomenda-se a consulta aos Cadernos Técnicos do Projeto de Conservação Preventiva de Bibliotecas e Arquivos – CPBA, de nºs 1 a 9: **Armazenagem e Manuseio**; nºs 10 a 12: **Procedimentos de Conservação**, 13 - **Manual de Pequenos Reparos em Livros**, nº 39 - **Preservação de fotografias: métodos básicos para salvaguardar suas coleções** e, nº 41 - **Indicações para o cuidado e a identificação da base de filmes fotográficos**. Disponíveis em: <<http://siarq49.siarq.unicamp.br/cpba>>.

devido frequência de uso, estado de conservação, ou alto valor intrínseco⁷ com necessidade de incremento de sua segurança, sem entretanto descontextualizá-los do conjunto a que pertencem.

8. CAPTURA DIGITAL⁸ DA IMAGEM⁹

O processo de captura digital da imagem deverá ser realizado com o objetivo de garantir o máximo de fidelidade entre o representante digital gerado e o documento original, levando em consideração suas características físicas, estado de conservação e finalidade de uso do representante digital. Recomenda-se a digitalização das capas, contracapas e envoltórios, bem como de páginas sem impressão (frente e verso) especialmente quando contiverem sinalização gráfica de numeração e outras informações.

No processo de captura digital dos documentos arquivísticos para conversão em imagem, deve-se observar os parâmetros que possam significar riscos ao documento original, desde as condições de manuseio, a definição dos equipamentos de captura, o tipo de iluminação, o estado de conservação até o valor intrínseco do documento original.

Os metadados técnicos a respeito do ambiente tecnológico (do documento original, da captura digital, do formato de arquivo digital gerado) e as características físicas dos documentos originais devem ser registrados em planilha e sempre que possível, devem ser encapsulados ao próprio objeto digital ou armazenados em um banco de dados.

É necessário que os equipamentos utilizados possibilitem a captura digital de um documento arquivístico de forma a garantir a geração de um representante digital que reproduza, **no mínimo**, a mesma dimensão física e cores do original em **escala 1:1**, sem qualquer tipo de processamento posterior através de softwares de tratamento de imagem. (V. Tabela 1).

PARÂMETROS PARA A OBTENÇÃO DE QUALIDADE DA IMAGEM DIGITAL

A qualidade da imagem digital é o resultado dos seguintes fatores: da resolução óptica adotada no escaneamento¹⁰, da profundidade de bit, dos processos de interpolação (quando utilizados) e dos níveis de compressão, além das características dos próprios equipamentos e técnicas utilizadas nos procedimentos que resultam no objeto digital.

7 Valor intrínseco : “Valor que um documento possui em razão de seu conteúdo, das circunstâncias de sua produção, de suas assinaturas ou selo”. Cf. Dicionário brasileiro de terminologia arquivística. Rio de Janeiro: Arquivo Nacional, 2005. 232 p. – Publicações Técnicas; n° 51.

8 V. nota 4.

9 Recomendamos a consulta aos Cadernos Técnicos do CBPA, de n° 51: **Requisitos de resolução digital para textos: métodos para o estabelecimento de critérios de qualidade de imagem**; e de n° 44, **O básico sobre o processo de digitalizar imagens**. Disponíveis em: <<http://siarq49.siarq.unicamp.br/cpba>>.

Ver também a recomendação do National Archives and Records Administration (NARA): **Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images For the Following Record Types- Textual, Graphic Illustrations/Artwork/Originals, Maps, Plans, Oversized, Photographs, Aerial Photographs, and Objects/Artifacts** e as referências contidas no seu **Anexo II**.

10 Escaneamento: Nessa recomendação é o termo que designa a conversão em imagem em arquivo digital por meio de um utilizando-se dispositivo eletrônico.

A resolução linear é determinada pelo número de pixels¹¹ utilizados para apresentar a imagem, e expressa em pontos por polegada (dpi) ou pixels por polegada (ppi) da vertical e horizontal da imagem digital (eixo X,Y). Quanto maior o número de pixels utilizados no processo de captura digital de imagem, mais elevada será a resolução linear e, portanto, a possibilidade de representar a imagem original com a riqueza de detalhes do documento original.

A profundidade de bit (também chamada de resolução tonal, resolução de cor ou variação dinâmica), é uma medida do número de bits utilizados para definir cada pixel. Quanto maior o número de bits para compor cada pixel, maior será a escala de tonalidades de cinza (*greyscale*) – onde há um bit por pixel para as cores (modo de cores) a serem apresentadas. Quando só se utiliza um bit por pixel denominamos de bitonal, ou seja, há apenas o preto e o branco.

A compressão de formato de imagem digital é um recurso amplamente utilizado, tanto para armazenamento quanto para a transmissão de dados, e existem muitos tipos de formatos de arquivo digitais e de *software* de imagem permitem a sua compressão, o que os tornam menores em volume de bits. Existem formatos de compressão sem perda de qualidade aparente (*lossless*), ou de compressão com perdas (*lossy*). A compressão não deve afetar a qualidade da imagem digital em relação a sua fidelidade com o original na recomendada **escala 1:1**.

Após a captura digital devem ser realizadas as atividades de controle de qualidade, conforme será descrito mais adiante.

Importante ressaltar que o parâmetro de qualidade é o de **Resolução Óptica**, que é a capacidade de captura real da imagem, em quantidades de pontos, sem a utilização de recursos de interpolação, que resultam num aumento artificial da resolução. Não se deve utilizar filtros para a geração das matrizes digitais.

A utilização da interpolação¹² em imagens digitais consiste na adição, por meio de *software*, de novos *pixels*, a partir dos *pixels* existentes. Seu propósito é fazer com que uma imagem digital pareça sido capturada originalmente com maior resolução. É para uso, por exemplo, em imagens *pequenas*, como *thumbnails* em sítios da internet é um recurso que não pode ser utilizado para a geração de matrizes digitais.

Outro fator a ser considerado são as condições de iluminação direta e indireta do local de trabalho. Deve-se obter um ambiente com controle das fontes de luminosidade, sejam os refletores utilizados para iluminar diretamente o documento seja as condições do local destinado a essa atividade (tipos e posicionamento), adequando-o aos requisitos técnicos necessários, como cor de paredes e piso, controle de aberturas (janelas e portas). O uso de roupas de cores fortes deve ser evitado no local onde é realizada a operação de captura digital. A calibração final das condições do local onde está equipamento e o acervo a ser digitalizado é feita utilizando-se um cartão de Referência de Branco (*White Reference*).

TIPOS DE EQUIPAMENTOS PARA CAPTURA DIGITAL DE IMAGEM

11 Pixel: é o menor ponto que forma uma imagem digital. Quanto maior for o número de *pixels*, melhor a resolução da imagem. Utilizado para estabelecer a resolução da imagem em formato digital e padrão de visualização em tela.

12 Interpolação: aumento artificial, por meio de software, da quantidade de pontos (*pixels*).

Atualmente existem disponíveis no mercado diversos tipos de equipamentos de captura digital para imagens, que se aplicam aos diversos tipos de documentos arquivísticos. A definição do equipamento de captura digital a ser utilizado só poderá ser realizada após o minucioso exame do suporte original, considerando suas características físicas e estado de conservação, de forma a garantir aos representantes digitais a melhor fidelidade visual em relação aos documentos originais, e sem comprometer seu estado de conservação.

Escâneres de mesa (*flat bed*)

Considerando a dimensão do item documental que não poderá exceder a área de escaneamento, são indicados para os documentos planos em folha simples e ampliações fotográficas contemporâneas em bom estado de conservação. Este tipo de equipamento não se aplica a documentos encadernados.

Escâneres planetários

Este tipo de equipamento utiliza uma unidade de captura semelhante a uma câmera fotográfica, uma **mesa de reprodução**¹³ que define a área de escaneamento e uma fonte de luz. São usados para a digitalização de documentos planos em folha simples, de documentos encadernados que necessitem de compensação de lombada, de forma a garantir a integridade física dos mesmos, bem como para os documentos fisicamente frágeis, já que não ocorre nenhuma forma de tração ou pressão mecânica sobre os documentos.

Câmeras digitais

O uso de câmeras digitais implica no uso de mesas de reprodução, para a garantia do paralelismo necessário à uma boa qualidade da imagem digital gerada, além de sistemas de iluminação artificial compatíveis, necessariamente com baixa intensidade de calor e o mínimo de tempo de exposição necessário para não comprometer o estado de conservação dos documentos arquivísticos originais, em especial os itens coloridos e as fotografias produzidas com processos fotográficos não contemporâneos, como daguerreótipos, albuminados e ferrótipos¹⁴.

Recomenda-se o uso de câmeras de médio e grande formato com *backs* digitais¹⁵ para geração de representantes digitais de alta qualidade, e para a captura digital de documentos em grandes formatos como mapas e plantas. Sempre que possível, deve-se privilegiar sistemas planetários de captura para evitar riscos de manuseio dos originais a serem digitalizados, principalmente quando se tratar de documentos frágeis e encadernados.

As câmeras digitais geram um arquivo digital denominado de RAW¹⁶, e que é um formato proprietário (cada fabricante tem o seu próprio formato que o faz dependente de *hardware* e *software* específicos). Entendemos que a utilização da câmera digital para

13 Há modelos de mesa de reprodução disponíveis que possuem um sistema de sucção que permite a reprodução de documentos de grandes formatos sem a formação de dobras e mossa na superfície do documento.

14 Processos fotográficos: Consultar os Cadernos Técnicos do CBPA de nº 39 - **Preservação de fotografias: métodos básicos de salvar suas coleções**, nº 40 - **Guia do Image Permanence Institute (IPI) para armazenamento de filmes de acetato** e nº 41 - **Indicações para o cuidado e a identificação da base de filmes fotográficos**.

15 É o dispositivo de captura digital (sensor - CCD) acoplado a uma câmera fotográfica dotada de conjunto óptico (lentes) convencional ou híbrido.

16 RAW. É o formato de arquivo digital das câmeras digitais e de escaners. São formatos proprietários. Em geral as imagens em RAW são convertidas para formatos como o TIF ou JPEG. Em fotografia digital são muitas vezes denominados de "negativos digitais" pois são o registro em arquivo digital. Ver nota 17.

reproduzir em imagens digitais documentos permanentes é direcionada à produção de um representante digital do documento original em outro suporte e não à produção de um original digital, e portanto não há necessidade obrigatória de conservar o formato RAW¹⁷ após a finalização do processo de captura digital e controle de qualidade e geração da imagem matriz e das imagens derivadas.

Equipamentos para digitalização de negativos e diapositivos fotográficos

Deve-se utilizar escâneres específicos para a captura deste tipo de documento, preferencialmente multiformato.¹⁸ Neste tipo de equipamento só podem ser utilizados negativos e diapositivos de suporte flexível e em bom estado de conservação.

Negativos e diapositivos de vidro, bem como negativos e diapositivos já em processo de deterioração não podem ser digitalizados neste tipo de equipamento devido ao risco causado pelo modo de operar seus dispositivos mecânicos e ópticos, devendo-se então utilizar um sistema de captura formado por câmeras digitais, mesas de reprodução e caixas de luz contínua ou com sistema de flash, como sistema de retro-iluminação. Ver Tabela 1.

Equipamentos para digitalização de microformas

Recomenda-se utilizar os escâneres específicos para a captura digital de diferentes microformas¹⁹, garantindo a melhor fidelidade em relação ao original e integridade física daqueles tipos de documentos. A qualidade das imagens obtidas poderá variar em função do estado de conservação dos filmes. Ver Tabela 1.

Escâneres de produção e alimentação automática

O incremento das redes de dados (internet), permitindo ampla disponibilização dos documentos em formato digital, tem levado a sociedade em geral a demandar que as organizações arquivísticas invistam em projetos de digitalização de massa (grande volume de itens) de seus acervos documentais, e nesse sentido, tem sido avaliada e testada a utilização de equipamento de captura digital com mecanismos de alimentação automática e maior velocidade de operação²⁰, embora algumas organizações arquivísticas ainda não aceitem a

17 A preservação de arquivos digitais em formato RAW tem sido abordada, por exemplo, com a sua conversão para o DNG (Adobe Digital Negative, criado pela @Adobe) ou para uma solução em desenvolvimento denominada de OpenRaw, a fim de evitar a manutenção de inúmeros e diversos formatos proprietários RAW. Sobre essa questão ver: *Camera Raw Formats (Group Description)* em <http://www.digitalpreservation.gov/formats/fdd/fdd000241.shtml> e Adobe Digital Negative (DNG) em <http://www.digitalpreservation.gov/formats/fdd/fdd000188.shtml> e OpenRaw em <http://www.openraw.org>.

18 Escâneres multiformato possuem dispositivos adaptadores que permitem o manuseio de formatos de filmes fotográficos em tamanhos e tipos variados, como tiras de 16mm, 35mm, cromos, etc.

19 Para microformas, ver o Manual do RLG para microfilmagem de arquivos. Projeto Conservação Preventiva em Bibliotecas e Arquivos: Arquivo Nacional. 2001. Caderno Técnico nº 53. Rio de Janeiro. Disponível em: <http://siarq49.siarq.unicamp.br/cpba>.

20 Digitization and Preservation of Archival Material – Principles and Dogmas. Lajos Körmeny - National Archives of Hungary. Disponível em: <http://www.liv.ac.uk/researchintelligence/issue20/pdf/ri20.pdf>

livre utilização desses equipamentos em documentos permanentes²¹.

A opção em empregar estes equipamentos para captura digital de documentos com sistemas de alimentação automática, também conhecidos como escâneres de produção, deve ser meticulosamente avaliada, devido a risco potencial de danos físicos e de redução da longevidade de documentos originais, em virtude do modo de operar de seus dispositivos mecânicos e ópticos, uma vez que é irreversível o modo de operação no momento quando estão em contato com o documento original, e a sua exposição a luminosidade artificial intensa e ao calor emanado pelo equipamento.

Uma vez realizada a opção pelo uso desses equipamentos quando do planejamento de projeto de digitalização de massa, e restringindo seu uso em acervo documental que se apresente em excelente estado de conservação (sem danos, rasgos), com baixo valor intrínseco, elevada demanda de acesso, e com alto índice de homogeneidade de sua constituição físico-química, dimensões, tipo de gramatura do papel (por exemplo: papel-cópia, papel-carbono, papel-de-arroz em geral são sempre muito frágeis).

Nesse sentido, na avaliação dos equipamentos para captura digital a serem utilizados em projetos de digitalização de massa, aqueles devem ser adequados ao tipo de documento original, não empregando-os naqueles com grande valor intrínseco²² atribuído ao documento arquivístico original (valor artístico, raridade, valor histórico), levando em consideração o seu estado de conservação, as suas características físico-químicas (dimensão, gramatura do papel, tipo de papel, tipo de tinta). Não é possível a utilização destes equipamentos para documentos arquivísticos que possuam elementos não planos como selos de cera, selos de papel, marcas em relevo, etc.

Todos os documentos a serem digitalizados neste tipo de equipamento deverão obrigatoriamente passar por um processo de análise rigorosa de sua estrutura física, seu estado de conservação, bem como a retirada de sujidades e objetos como clips, grampos, fitas adesivas e assemelhados.

Documentos fotográficos (em película, papel fotográfico e assemelhados)²³ e material de arte, não podem ser digitalizados nestes tipos de equipamentos com alimentação automática, independente de seu tipo e estado de conservação.

Existem, no mercado diversos tipos de equipamentos que exercem diferentes tipos de tração e pressão mecânica sobre os documentos (sistema de alimentação), e intensidade de luminosidade e calor. Neste sentido, deve-se optar por equipamentos que exerçam tração linear sobre o documento e não devem ser utilizados aqueles dotados de tambores (*drum scanners*).

Outros elementos a considerar:

21 a) "*The unit/partner/contractor shall not use automatic feed devices, drum scanners or other machines that require archival materials to be fed into rollers or wrapped around rollers, that place excessive pressure on archival materials, or require the document to be taped to a cylinder. Motorized transport is acceptable when scanning microfilm*". In Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Files, Raster Images For the Following Record Types- Textual, Graphic Illustrations, Artwork, Originals, Maps, Plans, Oversized, Photographs, Aerial Photographs, and Objects, Artifacts.. Written by Steven Puglia, Jeffrey Reed, and Erin Rhodes. Maryland, 2004. pag.179. [National Archives and Records Administration] - Disponível em: <<http://www.archives.gov/preservation/technical/guidelines.html>> Acesso: Dez. 2009.

22 Valor intrínseco : Valor que um documento possui em razão de seu conteúdo, das circunstâncias de sua produção, de suas assinaturas ou selo. Dicionário Brasileiro de Terminologia Arquivística. Rio de Janeiro : Arquivo Nacional, 2005. pg. 170.

23 Aqui deve-se distinguir acervo/documento fotográfico daquelas imagens que são simplesmente impressas em papel ou material assemelhado.

- Assistência de técnico/especialista em conservação de documentos no planejamento e execução do projeto de digitalização;
- Sempre utilizar os parâmetros técnicos de utilização do equipamento indicados pelo fabricante e fornecedor, adequando o funcionamento do equipamento as características do documento original a ser digitalizado;
- Adquirir/utilizar equipamentos com que sejam capazes de manusear papéis com gramaturas e espessuras muito variadas, e dotados de sensores que parem o funcionamento quando da ocorrência de material fora da especificação desejada (*built-in sensors*).
- Fazer testes antes da utilização;
- Operação com pessoal técnico que conheça o equipamento;
- Agrupar de forma homogênea o acervo original (tamanho/dimensões, tipo de papel, gramatura, estado de conservação, tipo de tinta).

SINALÉTICA²⁴

Em imagem digital as sinaléticas básicas são as seguintes:

- Escala de cinza;
- Escala de cores;
- Escalas e sinalização de dimensões.

Não se deve inserir na imagem digital uma sinalética tão próxima à imagem. Deve-se manter uma pequena margem ao redor do documento que permita um recorte posterior para a produção de formatos de arquivo digital derivados para acesso.

FORMATOS DOS ARQUIVOS DOS REPRESENTANTES DIGITAIS

24 Sobre sinaléticas em imagens digitais ver: **Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images For the Following Record Types- Textual, Graphic Illustrations/Artwork/Originals, Maps, Plans, Oversized, Photographs, Aerial Photographs, and Objects/Artifacts.- U.S. National Archives and Records Administration (NARA).** Disponível em: <http://www.archives.gov/preservation/technical/guidelines.html>. Acesso em abril de 2009.

Para a geração de matrizes e derivadas em formatos de arquivo digitais, recomenda-se sempre a adoção dos formatos abertos (*open sources*)^{25, 26} por permitirem melhores condições de acesso e preservação em longo prazo, e uma menor dependência de *software* e *hardware*.²⁷

O formato mais utilizado para os representantes digitais matrizes é o formato **TIFF** (**T**agged **I**mage **F**ile **F**ormat)²⁸, que apresenta elevada definição de cores sendo amplamente conhecido e utilizado para o intercâmbio de representantes digitais entre as diversas plataformas de tecnologia da informação existentes.

Também pode ser apreciado o uso de outros dois formatos digitais: o formato Portable Network Graphics - PNG²⁹ e o formato JPEG 2000³⁰.

O formato de arquivo digital Portable Network Graphics – PNG, surgiu inicialmente para substituir o formato GIF, e tem encontrado aceitação para a geração de matrizes digitais e tem características semelhantes ao TIFF. Uma de suas vantagens é utilizar uma compressão sem perdas, além, de ser um formato padronizado pela *International Standard Organization* como ISO/IEC 15948:2003. Entretanto, é mais limitado na inserção de metadados embutidos.

E o formato de arquivo digital JPEG 2000, tem sido apreciado para a geração de matrizes quando os originais em outro formato continuam a serem preservados, mas tem atualmente limitações em navegação WEB, devendo ser gerada uma imagem derivada de acesso em JPEG. Pode ser configurado para fazer a compressão sem perdas. Em relação ao PNG, esse formato permite embutir mais metadados. É um formato padronizado pela *International Standard Organization* como ISO/IEC 15444-1:2000.

O processo de captura digital, a partir dos documentos originais, deverá, necessariamente, gerar representantes digitais de alta e baixa resoluções, denominados respectivamente, **Matrizes e derivadas**:

1) MATRIZ DIGITAL – MD

25 “Formato de arquivo: Especificação de regras e padrões descritos formalmente para interpretação dos bits constituintes de um arquivo digital. Pode ser: 1. aberto – quando as especificações são públicas (p. ex.: XML, HTML, ODF e RTF); 2. fechado – quando as especificações não são divulgadas pelo proprietário (p. ex.: doc); 3. proprietário – quando as especificações são definidas por uma empresa que mantém seus direitos, sendo seu uso gratuito ou não (p. ex.: pdf, [...], doc e gif)”; 4. padronizado – quando as especificações são produzidas por um organismo de normalização, sendo os formatos abertos e não proprietários (p. ex.: XML). (I) Format; (F) format; (E) formato. In: Conselho Nacional de Arquivos. CONARQ. Câmara Técnica de Documentos Eletrônicos – CTDE. Glossário, 2008 - Disponível em: <http://www.documentoseletronicos.arquivonacional.gov.br/media/publicacoes/glossario/2009glossario_v5.0_final.pdf> Acesso: Dez. 2009.

26 Interpares2 Project (2007) General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation. Ver em: <http://www.interpares.org/display_file.cfm?doc=ip2_file_formats%28complete%29.pdf> Acesso: Dez. 2009.

27 Para conhecer uma abordagem atualizada sobre formatos de matrizes digitais, recomendamos o estudo da National Library of Netherlands: *Alternative File Formats for Storing Master Images of Digitisation Projects*. Disponível em: <http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicaties/Alternative%20File%20Formats%20for%20Storing%20Masters%20%201.pdf> Acesso: Abr. 2009.

28 Na data de publicação dessa recomendação, a última versão desse formato digital de imagem é o TIFF 6.0. Para efeito de atualização, recomendamos, quando da realização de projeto de digitalização. Disponível em: <<http://partners.adobe.com/public/developer/tiff/index.html>>. Acesso em abril de 2009.

29 PNG – <<http://www.digitalpreservation.gov/formats/fdd/fdd000138.shtml>>. Acesso: Dez. 2009

30 JPEG 2000 – <<http://www.digitalpreservation.gov/formats/fdd/fdd000138.shtml>>. Acesso. Dez. 2009

Os representantes digitais denominados como tal deverão ter alta qualidade de captura (resolução óptica em dpi e profundidade de bit) e ser armazenados e gerenciados por profissionais altamente qualificados em Tecnologia da Informação; o acesso deverá ser restrito e sob nenhuma hipótese autorizado a usuários não credenciados. O armazenamento desta matriz deverá ser feito em ambiente altamente protegido e fora dos sistemas e redes de dados para acesso remoto.

Recomenda-se a adoção de resolução óptica, profundidades de bit e modo de cor de acordo com a Tabela 1 para a obtenção de um MD com capacidade de reproduzir o original em escala 1:1.

Maiores valores de resolução linear em dpi podem ser considerados em razão das características do documento original, e da opção em se ter os representantes digitais com elevado nível de detalhamento. A adoção de resolução acima de 300 dpi implicará, portanto, na criação de matrizes digitais com maior tamanho em bits e necessidade de previsão de maior espaço de armazenamento.

Além da própria MD, deve-se criar duas cópias de segurança sendo uma a ser armazenada em um servidor com sistema de espelhamento do tipo RAID³¹ com acesso restrito e uma em fitas magnéticas do tipo *Linear Tape Open*.

2) MATRIZ DIGITAL COM PROCESSAMENTO DE IMAGEM – MDPI (opcional)

Recomendamos, opcionalmente, a geração do que denominamos de **Matriz Digital com Processamento de Imagem – MDPI**, que consiste na geração, com o auxílio de processamento de imagem, em formato TIFF, de uma matriz de alta resolução, com **compressão sem perda de qualidade aparente**, mas que permita uma melhor visualização e acesso, com a ausência de margens pretas e das sinaléticas presentes na MD. No entanto, deve-se evitar interferências estéticas e o uso de filtros e outras ferramentas que alterem ou distorçam o representante digital em termos de fidelidade visual ao documento original. A integridade e autenticidade da **MDPI** é dada pela geração e manutenção da **MD** da qual esta deriva.

31 R.A.I.D – Acrônimo de Redundant Array of Independent Discs ou Redundant Array of Inexpensive Discs (as vezes pode-se grafar como *Drives* ao invés de *Discs*). “A idéia básica por trás do RAID é combinar diversos discos pequenos e de custo baixo em um conjunto, para atingir objetivos de desempenho ou redundância inatingíveis com um disco grande e de custo alto. Este conjunto de discos aparece para o computador como uma única unidade ou disco de armazenamento lógico. [...]. O conceito fundamental do RAID é que os dados podem ser distribuídos ao longo de cada disco do conjunto de maneira consistente. Para fazer isso, primeiramente os dados precisam ser quebrados em *pedaços* de tamanho consistente [...] Cada pedaço é então gravado em um disco rígido no RAID, conforme o nível do RAID usado. Quando os dados tiverem que ser acessados, o processo é revertido, dando a impressão de que os discos múltiplos são um disco grande”. Disponível em: <http://web.mit.edu/rhel-doc/3/rhel-sag-pt_br-3/ch-raid-intro.html> Acesso em abril de 2009.

Essas cópias devem ser armazenadas em locais diferentes. A partir da **MDPI** é que serão gerados posteriormente os outros formatos de representantes digitais derivadas, e de acordo com a necessidade dos usuários finais³².

3) FORMATOS DE ACESSO

São os representantes digitais de acesso para visualização em tela, de navegação (*thumbnails*), de impressão e para *download* e que são gerados a partir das **MD** para diferentes usos e, sempre que possível, de acordo com a demanda dos usuários finais.

DERIVADAS DE ACESSO – DA

A partir da Matriz Digital ou da Matriz Digital com Processamento de Imagem (caso seja utilizada), serão criados um ou mais representantes digitais, que denominamos nessa recomendação como **Derivadas de Acesso**, com compressão e menor resolução linear, facilitando o seu acesso, disseminação e uso.

Recomenda-se que as derivadas de acesso destinadas aos usuários finais - para visualização em tela, impressão, *download*³³ ou cópia por demanda³⁴, sejam disponibilizadas preferencialmente em formato aberto, e em variados tamanhos de acordo com a banda utilizada por esse(s) mesmo(s) usuário(s) (conexão discada a 56 kbps, bandas largas de 256 kbps a 12 mbps, ou maiores).

As derivadas de acesso podem receber tratamento de imagem a fim de permitir melhor visualização ou impressão. Entretanto, critérios éticos devem pautar esse tipo de intervenção para que elas não se tornem dissociadas e não representem corretamente o documento original que as gerou. Sempre que possível, deverão ser utilizados preferencialmente formatos abertos para a geração dessas derivadas, e recomendamos os formatos **JPEG**³⁵ e **PNG**. No entanto, pode ser necessária a utilização de outros formatos de arquivo digitais, até mesmo proprietários, em virtude de concepção de acesso por parte dos provedores e custodiadores dos representantes digitais.

32 Para saber mais sobre essa modalidade e utilização de matriz digital com processamento de imagem: O National Archives (NARA) sugere a criação de *Production File Masters*, e a Índia os denominados *Clean Master* como *Masters* intermediários entre as Matrizes e os formatos de arquivo digital derivadas. Ver no *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images For the Following Record Types- Textual, Graphic Illustrations/Artwork/Originals, Maps, Plans, Oversized, Photographs, Aerial Photographs, and Objects/Artifacts.- U.S. National Archives and Records Administration (NARA). APPENDIX A: Digitizing for Preservation vs. Production Masters*, p.66: Disponível em: <<http://www.archives.gov/preservation/technical/guidelines.pdf>> Sugerimos também consulta ao *Guidelines for Digitization of Manuscripts*. National Mission for Manuscripts, 2005. Índia. *Output Specification*, p.15. Disponível em: <<http://namami.nic.in/DigiStds.htm>>. Acesso em abril de 2009.

33 *Download* é o recurso de disponibilizar arquivos em formato digital por meio da internet ou intranet, permitindo a sua transferência para um dispositivos de armazenamento remoto.

34 Cópia por demanda. Essa é uma modalidade de reprodução que pode ser realizada a partir de uma gama de formatos de arquivos digitais oferecidos e formas de entrega por meio de redes de dados, ou em mídias de armazenamento. A partir de solicitação individual direta a organização (pedido de serviço) ou por meio automatizado, diferenciados formatos de arquivo digitais podem ser gerados e disponibilizados ao usuário final, como por exemplo de um item ou vários itens documentais (variadas imagens) onde esses são convertidos em um único documento em formato como o PDF.

35 JPEG - Joint Photographic Experts Group. Formato digital de imagem. É um formato de arquivo digital com compressão com perdas (*lossy*), padronizado pela ISO, International Standard Organization – <http://www.iso.org>, de ampla aceitação e uso, o que o limita como melhor opção para matriz digital. Tem sido avaliado o uso do formato JPEG 2000, para a produção de matrizes digitais. Disponível em: <<http://www.jpeg.org/>>. Acesso em abril de 2009.

O formato de arquivo digital *Portable Document Format* - PDF ou PDF/A³⁶ também é recomendado, embora possua uma taxa de compressão menor. Esse formato digital permite dar acesso ao usuário final uma representação fiel do documento original, em um único arquivo digital, especialmente quando esse é formado por múltiplas páginas e contiver também imagens fixas.

No caso de documentos originais de grandes dimensões, podem ser utilizados formatos, com alta taxa de compressão e tecnologia *wavelet* que permite uma descompressão seletiva e a visualização de partes, orientadas por um navegador, facilitando a sua visualização pelos usuários finais, como por exemplo, o formato DJVU³⁷

DERIVADA DE NAVEGAÇÃO (*Thumbnail*) – DN

Para a visualização de imagem a título ilustrativo em sítios da internet, deve-se utilizar um representante digital com baixa resolução, do tipo *thumbnail*. Nessa modalidade os formatos de arquivo digitais mais comuns são o GIF, BMP, PNG e JPEG. Esse formato apresenta uma resolução final que não é recomendada para textos ou desenhos com linhas muito finas ou caracteres muito pequenos, tornando-os, por vezes, pouco legíveis.

GERAÇÃO DE MATRIZES DIGITAIS

Seguem abaixo, na tabela 1, as recomendações para a captura digital, de acordo com o tipo de documento original, indicando o tipo de reprodução (modo de cor), formato de arquivo digital e resolução mínima em dpi.

36 PDF - Portable Document Format. - Formato digital desenvolvido pela @Adobe em 1993, e em 2008 adotado como padrão ISO-3200, e “que permite capturar e visualizar informações consistentes - a partir de quase todo aplicativo ou sistema operacional - e compartilhá-las praticamente com qualquer pessoa, em qualquer lugar...”. Em 2005, a ISO adotou o padrão PDF/Archive - PDF/A - ISO 19005-1:2005, que assegura o acesso a longo prazo. Disponível em: <http://www.adobe.com/br/products/acrobat/adobepdf.html>

37 Ver em <http://djvu.org/>. Acesso: Dez. 2009

TABELA 1

Tipo de documento	Tipo de Reprodução	Formato de arquivo digital	Resolução mínima, modo de cor e observações
Textos impressos, sem ilustração, preto e branco. (*) sem manchas	Bitonal (**)	TIFF ³⁸ sem compressão	Resolução mínima de 300 dpi ³⁹ , escala 1:1, com margem preta de 0,2 cm ao redor do documento, 4 bits, modo bitonal(**)
Textos impressos, com ilustração e preto e branco.(*) com manchas	Tons de cinza (***)	TIFF sem compressão	Resolução mínima de 300 dpi, escala 1:1, com margem preta de 0,2 cm ao redor do documento, 8 bits, modo tons de cinza (***)
Textos impressos, com ilustração e cor	Cor	TIFF sem compressão	Resolução mínima de 300 dpi, escala 1:1, com margem preta de 0,2 cm ao redor do documento, 24 bits (8 bits por canal de cor), modo RGB (****)
Manuscritos sem a presença de cor	Tons de cinza	TIFF sem compressão	Resolução mínima de 300 dpi, escala 1:1, com margem preta de 0,2cm ao redor do documento, 8 bits, modo tons de cinza (***)
Manuscritos com a presença de cor	Cor	TIFF sem compressão	Resolução mínima de 300 dpi, escala 1:1, com margem preta de 0,2 cm ao redor do documento, 24 bits (8 bits por canal de cor), modo RGB (****)
Fotografias (Preto e Branco e Cor) (**)	Cor	TIFF sem compressão	Resolução mínima de 300 dpi, escala 1:1, com margem preta de 0,2 cm ao redor do documento, 24 bits (8 bits por canal de cor), modo RGB, com carta de cinza para ajuste de níveis (preferencialmente)
Negativos fotográficos e diapositivos (a)	Cor	TIFF sem compressão	Resolução mínima de 3000 dpi, 24 bits (8 bits por canal de cor), modo RGB (****)
Documentos cartográficos	Cor	TIFF sem compressão	Resolução mínima de 300 dpi, escala 1:1, com margem preta de 0,2cm ao redor do documento, 24 bits (8 bits por canal de cor), modo RGB, com carta de cinza para ajuste de níveis (preferencialmente) (****)
Plantas	Preto e branco	TIFF	Resolução mínima de 600 dpi, 8 bits, com possibilidade de modo tons de cinza (****)
Microfilmes e microfichas	Tons de cinza (***)	TIFF sem compressão	Resolução mínima de 300 dpi, 8 bits, modo tons de cinza (***)
Gravuras, cartazes e desenhos (Preto e Branco e Cor)	Cor	TIFF sem compressão	Resolução mínima de 300 dpi, escala 1:1, com margem preta de 0,2cm ao redor do documento, 24 bits (8 bits por canal de cor), modo RGB, com carta de cinza ou cores para ajuste de níveis (preferencialmente) (****)

(*) Sem manchas / com manchas.

(**) **Bi-tonal:** Captura de imagem onde não há gradação entre o claro e o escuro. Recomenda-se o seu uso somente para textos impressos e/ou datilografados monocromáticos e muito homogêneos, sem presença de manchas ou escurecimento do suporte original.

(***) **Tons de Cinza / Escala de cinza:** (Greyscale) - Recomenda-se o uso de escala de cinza para evitar que pequenas manchas interfiram na leitura final do representante digital. Da mesma forma para a digitalização de microformas, caso tenham as características assinaladas acima.

(****) **Modo de cor: RGB** (Red-Green-Blue). Padrão de cores. RGB é a abreviatura do sistema de cores aditivas formado por Vermelho (Red), Verde (Green) e Azul (Blue). Este sistema é constituído por projeções de luz como monitores de vídeo e projetores (*data displays*), em contraposição ao sistema substrativo, formado por impressões (CMYK). O padrão RGB é padrão para apresentação de cores na internet e seu uso é recomendado para documentos originalmente coloridos ou com informações relevantes em cor e fotografias de modo geral.

38 Avaliar de acordo com o documento original a utilização do formato de arquivo PNG ao invés de TIFF.

39 A resolução óptica mínima de de 300 dpi é a recomendável quando se deseja utilizar a tecnologia OCR (Optical Character Recognition).

METADADOS TÉCNICOS

Esta recomendação explicita um esquema mínimo de metadados técnicos referentes à captura digital, e não abrange aqueles referentes a recursos de catalogação, indexação, descrição ou de busca e pesquisa.

Os outros esquemas de metadados não serão tratados nesse documento, a saber: metadados descritivos, metadados administrativos (em grande parte contemplado pelo esquema de metadados técnicos aqui recomendado) e metadados estruturais.

Os metadados técnicos descrevem as características do representante digital no que diz respeito ao processo de captura digital, onde deve ser descrito e registrado o ambiente tecnológico (*software e hardware*), bem como algumas das características físicas do documento, tais como tipo e dimensão. Estes metadados devem, preferencialmente estarem inseridos ou encapsulados no representante digital e também descritos e registrados em planilhas em banco de dados.

A maior parte dos metadados discriminados abaixo são produzidos e inseridos automaticamente seja no próprio representante digital ou em planilha de banco de dados.

Tabela 2

Identificador do documento – ID do representante digital
Dimensão física do original (inserção manual ou por escala)
Código de referência do documento original
Data de criação
Responsável pela criação
Data de modificação
Responsável pela modificação
Sistema de iluminação, quando for o caso
Formato do arquivo
Dimensão em pixel: Largura X / Altura Y
Profundidade de cor (resolução tonal ou de cor)
Modo de Cores = Bitonal, escala de cinza (grayscale) e RGB
Resolução linear = Pixel ou ponto por polegada (dpi ou ppi)
Tamanho do arquivo
Perfil de Cor = Padrão ICC ⁴⁰
Software de captura
Software de processamento de imagem
Sistema operacional
Hash (checksum) da imagem ⁴¹

40 *International Color Consortium* - ICC. Especificações para sistemas de cores em formatos e plataformas abertas. Disponível em: <<http://www.color.org/index.xalter>>. Acesso em abril de 2009.

41 *Hashing, hash*. Algoritmo que mapeia uma sequência de bits (de um arquivo em formato digital), com a finalidade de realizar a sua verificação de integridade (*checksum*).

IDENTIFICADOR DO REPRESENTANTE DIGITAL

É indispensável o uso de um identificador do representante digital, que seja unívoco, persistente, permanente e que sempre permita a sua localização e sua identificação.

Deve-se utilizar um esquema para geração do identificador (*file name*) que não permita existir no acervo dois representantes digitais com identificadores iguais. Esse identificador não é necessariamente o mesmo que é utilizado na codificação de referência do documento ou nome do documento. Deve ser sucinto, permanente, consistentemente estruturado.

Abaixo recomendações básicas para a elaboração e uso de identificadores⁴²:

seja único
seja consistentemente estruturado
seja considerado o máximo de itens a serem escaneados
ao utilizar um esquema numérico, use zeros à esquerda para facilitar a composição do identificador
seja concebido de forma pouco complexa e não muito longo de modo a evitar erros humanos
use minúsculas e extensões de arquivo (ex. TIFF, JPEG)
uso de números e/ou letras, mas não se recomenda o uso de caracteres de símbolos e espaços que possam dificultar a interoperabilidade
metadados dos documentos inseridos nos nomes dos arquivos (como data da captura, número de página) também devem ser replicados em local próprio, para prover segurança e facilidades quando da movimentação da imagem para outros sistemas ou na necessidade de ser renomeado/identificado
sequenciamento de informações e divisões estruturais importantes de imagens multi-partidas devem ser explicitamente registradas nos metadados estruturais e não apenas embutidos nos nomes dos arquivos
embora não seja recomendado incorporar informação em demasia no nome do arquivo, uma certa quantidade de informação pode servir como metadado descritivo mínima para o documento/arquivo/imagem, e é uma alternativa econômica para o fornecimento de dados mais detalhados em outra parte
na eventualidade de se utilizar um nome temporário, recomenda-se a utilização de um esquema numérico simples, associando-o intelectualmente a um nome mais complexo por meio de um banco de dados

42 Baseado e adaptado de: **Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Files – Raster Images For the Following Record Types- Textual, Graphic Illustrations/Artwork/Originals, Maps, Plans, Oversized, Photographs, Aerial Photographs, and Objects/Artifacts.** p. 60. VI. STORAGE : File Naming. Disponível em: <<http://www.archives.gov/preservation/technical/guidelines.pdf>> Acesso em abril de 2009.

CONTROLE DE QUALIDADE NO PROCESSAMENTO TÉCNICO DE CAPTURA DIGITAL

O controle de qualidade no processamento técnico de captura digital de imagem, deve ser realizado pelos responsáveis da captura digital da imagem, um conjunto de procedimentos técnicos com o propósito de efetuar a verificação da fidelidade do representante digital em relação ao documento original, e se foram obtidas as características técnicas requeridas como resolução, modo de cor e registro de metadados técnicos.

Recomenda-se a amostragem quando forem grandes volumes (quantidade de itens individuais) e com características muito homogêneas.

Recomenda-se o exame individual nos casos em que os documentos possuam grande valor intrínseco.

Formas de obtenção e verificação da qualidade:

Na tela (monitor calibrado) – Os monitores devem ser calibrados e estarem bem posicionados no ambiente, levando-se em consideração eventuais reflexos na tela. A calibragem deve ser feita de acordo com as normas do fabricante do monitor e perfil de Cor pelo Padrão ICC⁴³.

Ver a imagem em escala 1:1 - 100 %.

Uso de escala para avaliar escala de cinza ou modo de cor.

Uso de escalas de resolução e histogramas para avaliar a resolução espacial e a reprodução tonal.

Por impressão

Exame pelo olho humano, de cópias impressas para avaliar a qualidade da captura digital comparando com o documento original.

Controle de qualidade do escâner e/ou da câmera digital

Seguir os manuais dos equipamentos. Utilizar as escalas de cinza e cores padrões para a calibração. Verificar em amostras em tela e por impressão, e por comparação com o original. Utilizar softwares específicos para calibração de escâneres. Observar o perfil de Cor pelo Padrão ICC⁴⁴. Cada equipamento deve ser calibrado e gerado o seu ajuste individual denominado de ICC Profile.

Calibração de Monitores

Recomenda-se o uso de espectrofotômetros para calibragem dos monitores e a criação de perfis de cor para que sejam equivalentes às visualizações no monitor e na impressão. Existem uma gama de softwares e equipamentos específicos para essa tarefa.

Validação e Conformidade

43 Criação de Perfil ICC - *International Color Consortium* - Especificações para sistemas de cores em formatos e plataformas abertas. Disponível em: <<http://www.color.org/index.xalter>>. Acesso em abril de 2009.

44 Criação de Perfil ICC – V. nota 43.

Verificar se os os formatos de arquivo digitais gerados (Matrizes e derivadas de acesso) estão de acordo com as especificações técnicas . Essas especificações e as diferentes versões de um formato sempre estão publicadas pelo responsável por sua manutenção

Recomendamos a utilização do **Modelo de Verificação da Qualidade do Representante Digital (anexo I)**.

GERENCIAMENTO DOS REPRESENTANTES DIGITAIS

Recomenda-se adquirir ou desenvolver uma solução de *software* para o gerenciamento dos representantes digitais, que deverão ficar armazenados num repositório (ver: **Aspectos gerais para segurança, armazenamento e preservação das imagens digitais**) e operado por um SGBD⁴⁵. Esses sistemas devem permitir a interoperabilidade com outros sistemas informatizados da organização visando ao seu acesso, integração, manutenção e segurança. O objetivo desses sistemas é fazer o gerenciamento técnico e administrativo dos representantes digitais, por meio do controle intelectual (lógico) e da integridade (física) desses representantes digitais, sejam as Matrizes ou as Derivadas. Esse sistema é distinto de um sistema eletrônico de descrição arquivística e acesso a usuários finais.

9. ACESSO

Os representantes digitais, que serão colocados em acesso à própria organização e aos usuários finais deverão estar à disposição por meio de algum sistema eletrônico de acesso: (intranet e internet), para pesquisa, busca e recuperação, acesso, visualização, e *download* (se couber).

Recomenda-se que os formatos digitais destinados para uso de usuários finais (*download* ou cópia por demanda), sejam preferencialmente em formato digital aberto, e disponibilizados em variados tamanhos de acordo com a banda utilizada por esse mesmo usuário (conexão discada a 56 kbps, bandas largas de 256 kbps a 12 mbps).

10. ASPECTOS GERAIS PARA SEGURANÇA, ARMAZENAMENTO E PRESERVAÇÃO DAS IMAGENS DIGITAIS

Com relação aos aspectos que envolvem a segurança, o armazenamento e a preservação de documentos digitais, que não são o escopo principal dessa recomendação, indicamos abaixo alguns procedimentos e cuidados.

Nesse sentido, para uma visão geral, para recomendamos a consulta em primeiro lugar às diretrizes gerais apontadas na Resolução nº 25, do CONARQ, de abril de 2006, *Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos e-*

45 SGBD – Sistema de Gerenciamento de Bases de dados - É um software ou conjunto deles, responsável pelo gerenciamento de uma base de dados. O seu objetivo é de servir a aplicação cliente, e gerenciar o acesso, manipulação e organização dos dados. O SGBD disponibiliza uma interface para que os seus clientes possam incluir, alterar ou consultar dados. Em bancos de dados relacionais a interface é constituída pelas APIs ou drivers do SGBD, que executam comandos na linguagem SQL.

ARQ-Brasil, em seus ítems: 6. Segurança, 7. Armazenamento e 8. Preservação⁴⁶, e nas *Directrices para la Preservación del Patrimonio Digital*. Biblioteca Nacional de Austrália/UNESCO⁴⁷.

ARMAZENAMENTO

Para o armazenamento de representantes digitais de acervos de grandes dimensões e que envolvam grande quantidade de dados e portanto maiores dimensões de volume em *bytes* recomenda-se preferencialmente a utilização de memórias secundárias e terciárias. O mercado apresenta soluções de hardware específicos para armazenamento de massa (*data storage / mass storage*).

MATRIZES DIGITAIS

- Discos rígidos – HD - (*hard disks*), preferencialmente em configuração RAID⁴⁸ que incrementam a segurança.
- Fitas magnéticas do tipo LTO (*Linear Tape-Open*), e similares, que permitem o armazenamento *off-line* de grandes volumes de documentos digitais com custo mais baixo que os Hds.

DERIVADAS DE ACESSO

Os representantes digitais para acesso pelo usuário final (internos ou externos), deverão estar disponíveis por meio de unidades de HD ou mesmo em *juke-box* de CDs/DVDs, concebidas especificamente para esse fim.

Quando se pretender disponibilizar os representantes digitais em sistemas de acesso externo (sítios da internet), esses devem estar armazenados em dispositivos próprios, como os citados acima, mas não interligados (*on-line*) ao sistema de armazenamento e segurança das Matrizes Digitais.

CÓPIAS DE SEGURANÇA

Devem ser criadas ao menos duas matrizes digitais de segurança, e uma matriz com processamento de imagem (se for gerada). Essas cópias devem estar armazenadas fora de linha (*off-line*), ou seja sem conexão física na rede de dados e recomenda-se que uma cópia esteja em armazenamento *off-site* (armazenada em condições adequadas em instalação diversa e afastada da organização detentora do acervo).

Recomenda-se a geração e manutenção também de cópias de segurança dos representantes digitais (derivadas) para acesso.

Cada tipo de cópia, matriz ou derivada deve ser armazenada em dispositivo específico.

Devem ser criadas cópias de segurança das planilhas dos metadados técnicos que estejam em banco de dados, bem como dos demais metadados estruturais, descritivos e administrativos.

46 Disponível em: <<http://www.conarq.arquivonacional.gov.br/Media/publicacoes/earqbrasilv1.pdf>> Acesso em abril de 2009.

47 Disponível em: <<http://unesdoc.unesco.org/images/0013/001300/130071s.pdf>> Acesso em abril de 2009.

48 R.A.I.D – Acrônimo de Redundant Array of Independent Discs ou Redundant Array of Inexpensive Discs. Disponível em: <http://web.mit.edu/rhel-doc/3/rhel-sag-pt_br-3/ch-raid-intro.html> Acesso em abril de 2009.

11. UTILIZAÇÃO DE SERVIÇOS TERCEIRIZADOS DE DIGITALIZAÇÃO E CORRELATOS

Entende-se serviços terceirizados como aqueles que a contratante não pode realizar por seus próprios meios, tais como: captura de imagens, produção de matrizes digitais e formatos derivadas, armazenamento e serviços de acesso remoto a usuários finais. Esses serviços podem ser prestados de forma onerosa ou não.

Cabe esclarecer que essas recomendações gerais deverão seguir as normas legais sobre contratação de produtos e serviços dos órgãos e entidades públicas em sua esfera de atuação.

Devido ao custo elevado de um programa de digitalização, os formatos digitais matrizes gerados e entregues pela contratada devem ser de propriedade plena e irrestrita da contratante, e a fim de permitir a preservação e o acesso a longo prazo dos representantes digitais, preferencialmente, devem ser adotados formatos digitais abertos e não proprietários.

Como a captura digital resulta sempre no manuseio de documentos arquivísticos originais, em diversos estágios de estado de conservação, e muitos com alto valor intrínseco, dá-se preferência a que o manuseio, preparo e captura dos documentos originais sejam realizados no local da contratante e permanência dos documentos originais somente pelo tempo necessário para a captura da imagem, e o acondicionamento e manuseio dos documentos de acordo com o seu tipo e estado de conservação⁴⁹.

Na eventualidade de utilização de terceirização de serviços de captura digital, armazenamento e acesso aos representantes digitais por meio de serviços contratados ou parcerias sem ônus, são arroladas abaixo as responsabilidades mínimas das partes:

Responsabilidades mínimas da Contratante:

- Apresentar previamente o acervo identificado;
- Apresentar o acervo documental higienizado, planejado, reparado e acondicionado;
- Estabelecer previamente os formatos digitais a serem gerados e entregues;
- Avaliar a qualidade da organização a ser contratada e dos serviços e produtos que oferece por meio de atestados de capacidade, qualidade e de comprovação de serviços prestados anteriormente em atividade congênera e outros tipos de avaliações técnicas;
- Supervisionar o serviço contratado durante o período de execução;
- Evitar a movimentação do acervo original para o local de captura digital, sempre que possível.
- Avaliar o serviço e produtos contratados no momento de entrega (V. Anexo I).

Responsabilidades básicas da Contratada:

49 Cf. U.S. NATIONAL ARCHIVES AND RECORDS ADMINISTRATION. *Preservation Guidelines for Vendors Handling Records and Historical Materials*. Disponível em: <<http://www.archives.gov/preservation/technical/vendor-training.html>>. Acesso em abril de 2009.

- Manusear o acervo conforme as recomendações deste documento e as da organização contratante;
- Oferecer as condições físicas e técnicas para o manuseio e segurança do acervo original;
- Com relação à intervenção física no documento original, esta só poderá ser realizada por pessoal especializado em conservação e com acompanhamento permanente da Contratante ou por pessoa ou organização qualificada e autorizada por ela.
- Utilizar equipamentos que não ofereçam risco evidente ou potencial ao acervo
- Fazer o controle de qualidade da imagem digital e dos metadados técnicos;
- Realizar a entrega dos formatos de arquivos digitais requeridos e nas mídias de armazenamento⁵⁰ indicadas pela Contratante.

50 Mídia de armazenamento. A mídia de armazenamento do formato digital para efeito de entrega pela contratada a contratante, é em geral diverso daquela utilizada para o armazenamento definitivo. Por exemplo, os representantes digitais podem ser entregues em mídia móvel (discos ópticos) e que serão transferidos para *Hard Disks* de alta capacidade. Essa forma de operação técnica é também necessária para incrementar a segurança, uma vez que recomenda-se uma quarentena e análise para a verificação de falhas, corrupção de dados e de riscos como vírus informáticos..

ANEXO I

MODELO DE VERIFICAÇÃO DA QUALIDADE DO REPRESENTANTE DIGITAL

Essa lista de verificação é para se obter um indicativo de qualidade da captura digital e da criação dos representantes digitais matrizes digital e dos derivados.⁵¹

1. A imagem tem o tamanho/resolução pretendido?
a. Matriz Digital
b. Matriz Digital com Processamento da Imagem (opcional)
c. Formatos de arquivo digitais de Acesso (Derivadas de Acesso)
d. Thumbnail (Derivadas de Navegação)
e. Outros formatos de arquivo digitais derivadas de acesso
2. O nome da imagem digital esta correto?
a. Matriz Digital
b. Matriz Digital com Processamento da Imagem (opcional).
c. Formatos de arquivo digitais de Acesso
d. <i>Thumbnail</i> (Derivadas de Navegação)
e. Outros formatos derivados de acesso
3. O Formato Digital da imagem está correto?
a. Matriz Digital
b. Matriz Digital com Processamento de Imagem (opcional)
c. Formatos de arquivos digitais de Acesso
d. <i>Thumbnail</i> (Derivadas de Navegação)
e. Outros formatos de arquivo digitais derivadas de acesso
4. Verificação de Qualidade Digital com comparação com o documento original
a. A imagem esta correta no modo de cor
b. Recorte correto
c. Sem rotação
d. Sem inversão
e. Sem inclinação

51 Adaptado de: *Guidelines for Digitization of Manuscripts*. National Mission for Manuscripts, 2005. Índia. Disponível em: <<http://namami.nic.in/DigiStds.htm>>. Acesso em abril de 2009.

f. Perda de nitidez/ excesso de nitidez
g. Presença de interferência em imagens com linhas - Padrão Moiré ⁵²
h. Não pixelado
i. Predominância de uma das cores da imagem
j. Histograma ⁵³
k. Sem alto contraste / sem alta luminosidade
l. Valores tonais desiguais / reflexos
n. Aspecto granulado
m. Observações gerais

⁵² Padrão Moiré: É um padrão de interferência em captura digital de imagem, quando surgem contrastes (conhecidos como tramas ou grades), que não existem na imagem original, devido a superposição de linhas em ângulos diferenciados, ou porque elas tem tamanhos diferentes. Para evitá-lo, pode-se alterar o ângulo do objeto original no instante da captura, ou por processos em softwares de imagem.

⁵³ Histograma da imagem digital é a representação gráfica das medições da luminosidade da imagem no momento da captura, numa escala de 256 tons de cinza.

ANEXO II REFERENCIAS

- ARQUIVO NACIONAL (Brasil). **Dicionário brasileiro de terminologia arquivística**. Rio de Janeiro, 2005. 232p. 30cm. (Publicações Técnicas; nº 51). Disponível em: http://www.portalan.arquivonacional.gov.br/Media/Dicion_Term_Arquiv.pdf Acesso: 08 Dez. 2009.
- CANADIAN HERITAGE . **CCO Best Practice Guide** : Digitization Projects and the Preservation of Digital Content. Disponível em: <http://www.patrimoinecanadien.gc.ca/pgm/pce-ccop/publctn/tech-rgrmnts/bst_prctc_dgt-eng.cfm>. Acesso: 24 Fev. 2009.
- COLORADO DIGITIZATION PROGRAM (Estados Unidos). BCR's CDP Digital Imaging Best Practices Working Group. **BCR's CDP - Digital Imaging - Best Practices**. Version 2.0. Colorado, 2008. Disponível em: <<http://www.bcr.org/cdp/best/digital-imaging-bp.pdf> >. Acesso: 24 Abr. 2009.
- CONSELHO NACIONAL DE ARQUIVOS (Brasil). Resolução nº 6, de 15 de maio de 1997. Dispõe sobre diretrizes quanto à terceirização de serviços arquivísticos públicos. Disponível em: <http://www.conarq.arquivonacional.gov.br/cgi/cgilua.exe/sys/start.htm> Acesso: 24 Abr. 2009
- _____. Resolução nº 25, 27 de abril de 2007. Dispõe sobre a adoção do Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos - e-ARQ Brasil pelos órgãos e entidades integrantes do Sistema de Arquivos-SINAR. Disponível em: <http://www.conarq.arquivonacional.gov.br/Media/publicacoes/earqbrasilv1.pdf> Acesso: 24 Abr. 2009
- CONWAY, Paul. **Preservação no Universo Digital**. Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos; Arquivo Nacional, 2001.(Caderno Técnico nº 52). Disponível em: <http://siarq49.siarq.unicamp.br/cpba> Acesso: Abr. 2009.
- ELKINGTON, Nancy E. **Manual do RLG para microfilmagem de arquivos**. Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos; Arquivo Nacional, 2001.(Caderno Técnico nº 53). Disponível em: <http://siarq49.siarq.unicamp.br/cpba> Acesso: Abr. 2009.
- FISCHER, Monique C.; ROBB Andrew. **Indicações para o cuidado e a identificação da base de filmes fotográficos**. Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos; Arquivo Nacional, 2001. (Caderno Técnico nº 41).Disponível em: <http://siarq49.siarq.unicamp.br/cpba> Acesso: Abr. 2009.
- INDIA. National Mission for Manuscripts. **Guidelines for Digitization of Manuscripts**. (Nova Deli), 2005. Disponível em: <http://namami.org/digitization.pdf> Acesso: 24 Fev. 2009.
- KENNEY, Anne; CHAPMAN, Stephen. **Requisitos de resolução digital para textos**: métodos para o estabelecimento de critérios de qualidade de imagem. Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos; Arquivo Nacional, 2001. (Caderno Técnico nº 51).
- LIBRARY OF CONGRESS (Estados Unidos) .**Digital Preservation**. Disponível em: <http://www.digitalpreservation.gov/formats/fdd/descriptions.shtml> Acesso: 24 Abr. 2009
- _____.**Formats, Evaluation Factors, and Relationships**. Disponível em: <http://www.digitalpreservation.gov/formats/intro/intro.shtml> Acesso: 24 Abr. 2009
- MILEVSKY, Robert J. **Manual de Pequenos Reparos em Livros**. Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos; Arquivo Nacional, 2001. (Caderno Técnico nº 13).
- MUSTARDO, Peter; KENNEDY, Nora. **Preservação de fotografias**: métodos básicos de salvaguardar suas coleções. Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos; Arquivo Nacional, 2001.(Caderno Técnico nº 39).
- NATIONAL ARCHIVES AND RECORDS ADMINISTRATION (Estados Unidos). **Electronic Access Project Scanning and File Format Matrix**. Maryland, 1998. Disponível em: <http://www.archives.gov/preservation/technical/guidelines-matrix.pdf> Acesso: 24 Abr. 2009

- _____ **Frequently Asked Questions About Optical Media.** Disponível em: <http://www.archives.gov/records-mgmt/faqs/optical.html> Acesso: 24 Abr. 2009
- _____ **Preservation Guidelines for Vendors Handling Records and Historical Materials.** Maryland, 1998. Disponível em: <<http://www.archives.gov/preservation/technical/vendor-training.html>>. Acesso: 24 Abr. 2009.
- _____ **Technical Guidelines for Digitizing Archival Materials for Electronic Access:** Creation of Production Files, Raster Images For the Following Record Types- Textual, Graphic Illustrations, Artwork, Originals, Maps, Plans, Oversized, Photographs, Aerial Photographs, and Objects, Artifacts.. Written by Steven Puglia, Jeffrey Reed, and Erin Rhodes. Maryland, 2004. Disponível em: <http://www.archives.gov/preservation/technical/guidelines.html> Acesso: 24 Abr. 2009.
- _____ The Technology Research Staff the National Archives at College Park. **Digital-Imaging and Optical Digital Data Disk Storage Systems:** Long-Term Access Strategies for Federal Agencies. Maryland, 1994. (Technical Information Paper n°. 12). Disponível em: < <http://www.archives.gov/preservation/technical/imaging-storage-report.html#s6> >. Acesso: 24 Abr. 2009
- NATIONAL LIBRARY NEW ZEALAND. **Introduction to Digitisation.** Disponível em: <http://www.natlib.govt.nz/services/get-advice/digital-libraries/introduction-to-digitisation> Acesso: 24 Abr. 2009
- _____ ALEXANDER TURNBULL LIBRARY. **Digitisation Guidelines - Specifications for Imaging:** copying and digital services. Disponível em: <http://www.natlib.govt.nz/catalogues/library-documents/digitisation-guidelines/> Acesso: 24 Abr. 2009
- ONLINE COMPUTER LIBRARY CENTER (Estados Unidos). **Digital Archive™ :** Secure, managed storage for digital preservation. Disponível em: <http://www.oclc.org/us/en/digitalarchive/default.htm> Acesso: 24 Fev. 2009.
- OGDEN, Sherelyn (ed). **Armazenagem e Manuseio.** Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos; Arquivo Nacional, 2001. (Caderno Técnico nº 1 a 9). Disponível em: <http://siarq49.siarq.unicamp.br/cpba> Acesso: Abr. 2009.
- _____. **Procedimentos de Conservação.** Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos; Arquivo Nacional, 2001. (Caderno Técnico nº 10 a 12). Disponível em: <http://siarq49.siarq.unicamp.br/cpba> Acesso: Abr. 2009.
- REILLY, James M. Guia do Image Permanence Institute (IPI) para armazenamento de filmes de acetato. Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos; Arquivo Nacional, 2001.(Caderno Técnico nº 40). Disponível em: <http://siarq49.siarq.unicamp.br/cpba>. Acesso: Abr. 2009.
- UNESCO - **Directrices para la Preservación del Patrimonio Digital.** Preparado por la Biblioteca Nacional de Australia. División de la Sociedad de la Información. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. Disponível em: <http://unesdoc.unesco.org/images/0013/001300/130071s.pdf>. Acesso: 24 Fev. 2009
- WATERS, Donald J. **Do microfilme à imagem digital.** Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos; Arquivo Nacional, 2001. (Caderno Técnico nº 49).Disponível em: <http://siarq49.siarq.unicamp.br/cpba> Acesso: Dez. 2009.
- WILLIS, Don. **Uma abordagem de sistemas híbridos para a preservação de materiais impressos.** Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos, Arquivo Nacional, 2001. (Caderno Técnico nº 50). Disponível em: <http://siarq49.siarq.unicamp.br/cpba> Acesso: Abr. 2009.