

A digitalização de acervos

Primeiros passos

Já em 2004 o IEB iniciava a estruturação do setor de digitalização, contando, naquele momento, com um Scanner planetário Zeutschel OS8000, duas scanner HP Scanjet II e um digitalizador de microfilmes Kodak 2400DSV alocados na Seção de Processamento de Imagem e Reprografia (ScProire). Hoje, em 2010, o Instituto ampliou seu parque de máquinas, atualizou os scanners de mesa e montou uma estrutura de fotografia digital que conta com duas câmeras Nikon e uma Hasselblad H2, possibilitando não só uma produção regular na faixa de 100 mil imagens/ano, mas também um salto na sua qualidade, o que deu ao Instituto a possibilidade de produzir imagens de qualidade para originais de grande formato e de suportes variados. Todavia, a estruturação de um acervo digital vai além da compra de equipamentos. Assim, a proposta desse texto é descrever como esse processo se deu no IEB, que hoje se consolida como referência dentro e fora da Universidade.

A definição de padrões para digitalização

Atualmente, a discussão sobre padrões de digitalização já se encontra bem difundida e, apesar da ausência de consenso, existem experiências mais adiantadas que auxiliam na definição de parâmetros e escolhas.

O primeiro passo no processo de montagem de um centro de digitalização passa necessariamente pela adoção de padrões para a captura das imagens, para a disponibilização e a distribuição, e para a definição de cópias de segurança para os arquivos digitais. Essa determinação de padrões é indispensável e deve anteceder o processo de digitalização. Todavia, é necessário ter a consciência de que, no “mundo digital”, qualquer padrão é efêmero e deve ser reavaliado constantemente, em função da velocidade de atualização das tecnologias envolvidas.

Ao tratarmos de imagens digitais, estamos em uma área que, nos últimos anos, estendeu muito seus horizontes. Para se ter idéia, levando em conta somente a resolução das imagens, o que há cinco anos poderia ser obtido apenas com equipamentos profissionais de alto custo, hoje pode ser produzido em simples aparelhos celulares. Desta forma, para falarmos de padrões de digitalização temos de ter em mente que:

1. Não existem padrões nem produtos definitivos; todo e qualquer padrão deve ser constantemente reavaliado e toda produção digital (imagens, *softwares*, sistemas de busca, etc.) tendem a ser superados.
2. Nenhum padrão de armazenamento tem garantia de perenidade; tanto o formato quanto a mídia em que o arquivo está gravado tendem a se tornarem obsoletos.
3. Os padrões atuais de resolução são apenas referência ao disponível na época; a alta-resolução de hoje tende a ser a baixa-resolução de amanhã.

Essa constante mutação não diminui a importância da definição dos padrões, muito pelo contrário: é a existência de padrões pré-estabelecidos que permite uma migração segura. O fundamental é reavaliar regularmente os padrões adotados em todas as etapas do processo.

Padrões de captura na geração de matrizes digitais

Entendemos como matriz digital aquela imagem que, dentro dos limites técnicos impostos, reproduz o original da melhor forma possível. Sua produção segue padrões pré-estabelecidos para garantir a fidelidade. É importante lembrar que ela é produzida com o intuito de atender ao maior número de demandas possíveis, sendo que, a partir delas, são gerados os subprodutos para atender necessidades específicas.

Resolução de captura

Ao falarmos de padrões de captura, a primeira idéia que vem à mente da maioria das pessoas está ligada a sua resolução, porém este é apenas um dos vários itens a serem considerados.

Muito se fala em “300 dpi em tamanho original”, o que guarda sua validade, mas seu uso generalizado sem a real compreensão do sentido gera mais problemas que soluções. O olho humano é capaz, em média, de reconhecer, na distância de leitura, algo em torno de 280 pontos por polegada (dpi). Acima disso, vemos apenas como uma linha contínua. Assim, algo impresso a 300 dpi será percebido pelo observador como contínuo, sem que seja capaz de reconhecer os pontos que compõem a imagem. Temos então o primeiro problema: *dpi* é uma medida para a *saída* (impressão) e não para a captura da imagem, não fazendo sentido em relação ao tamanho do original. Para ficar claro: supondo a captura de uma pintura, originalmente com 1 m de largura, que será reproduzida em um catálogo, com metade do seu tamanho, precisaremos ter os 300 pontos por polegada em 50 cm e não em 1 m. O mesmo se aplica ao caso inverso: se a mesma pintura for reproduzida em um *banner* com 2 m de largura, os 300 dpi deverão aparecer nesta última medida.

Para a resolução de captura de imagem, utiliza-se uma outra medida, hoje popularizada pelas câmeras digitais, que representa o número de pontos que compõem a *imagem*: o megapixel. Apenas para critério de comparação, para se imprimir uma página tamanho A4 a 300 dpi é necessário que a imagem capturada tenha algo entorno de 8 milhões de pontos, ou seja, 8 megapixels.

Considerando que nem sempre se tem claramente no momento da digitalização qual será o tamanho da saída e nem qual a qualidade desta saída, via de regra, tende-se a optar pela solução aparentemente simples de “digitalizar na maior resolução possível”, o que, à primeira vista, pode parecer dar conta de todos os problemas. Infelizmente, a realidade não é bem esta, pois, quanto maior a resolução de captura, maior o tamanho do arquivo digital e, conseqüentemente, do espaço de armazenamento, do tempo para as gravações etc., tornando mais complexo e dispendioso todo o processo. Ainda assim, poderia-se pensar “teremos um maior custo, porém teremos uma ‘digitalização definitiva’!”, mas, como já dissemos anteriormente, não existe produto definitivo.

Diante deste quadro, no IEB estabelecemos a política de definir a resolução máxima a ser utilizada para cada tipo de digitalização a partir da análise de diferentes fatores como:

1. Aplicação imediata: em certos casos, o próprio uso previsto para determinada imagem já define será sua digitalização na qualidade máxima disponível.
2. Potencialidade de uso desta imagem: um texto impresso, cuja provável utilização seja a recuperação do conteúdo escrito, só precisa da resolução necessária para sua leitura e eventual impressão em dispositivos domésticos; já um manuscrito pode demandar maior capacidade de observação de detalhes para análises paleográficas, exigindo uma maior resolução para permitir sua ampliação.

3. Nível de detalhamento: imagens (mapas, pinturas, plantas arquitetônicas etc.) podem conter detalhes que precisem ser analisados separadamente, sendo necessário, portanto, uma captura que possibilite ampliação.
4. Dificuldade de acesso ao original e estado de conservação: as condições de acesso e manuseio do original (como estado de deterioração ou pertencimento a outra instituição) podem inviabilizar uma nova digitalização, justificando que esta seja feita na máxima resolução possível.

Consistência de cor

Algumas vezes, a decisão da forma de digitalização não está vinculada diretamente à resolução, mas sim à capacidade do equipamento em reproduzir com maior “fidelidade” as cores do documento original. Primeiro, é preciso lembrar que a cor é resultado de estímulos ao nervo ótico causado pela diferença na reflexão da luz, ou seja, além de depender da iluminação no momento da observação, é praticamente uma sensação do observador. Por isso, preferimos a idéia de *consistência de cor*, a qual consiste em possibilitar que, através de uma tabela de referência presente na imagem, esta possa ser reconstituída com uma cor semelhante à do documento, preservando assim as características originais – o que, em alguns casos, pode ser fundamental para a sua validade.

Neste processo, são utilizadas as tabelas de cinza (cartões impressos e certificados, com quadros brancos, pretos e cinzas) que servem de referência para a calibração das cores. Algumas dessas tabelas possuem escala métrica, fornecendo adicionalmente a referência do tamanho do original, permitindo sua recuperação pela comparação.

Integralidade do documento

Para que a reprodução resultante de digitalização de uma documentação histórica tenha validade para um pesquisador, precisa-se ter a certeza de sua integralidade. Para isso, alguns fatores são fundamentais.

1. Totalidade da imagem: ao digitalizar um documento, deve-se ter o cuidado de manter os limites externos separados do limite da área digitalizada por uma margem de segurança, de forma que o pesquisador possa ter a certeza de sua totalidade.
2. Sem intervenções na imagem: apesar da grande tentação de corrigir pequenos defeitos e imperfeições, adotamos como critério não interferir na imagem, para que o pesquisador tenha, dentro do possível, uma cópia fiel do original.
3. Registro do verso e páginas em branco: ao suprimir um verso ou uma página em branco, dá-se margem a dúvidas sobre sua integralidade, por isso, sempre que viável, procuramos registrá-los.

Cópias de segurança

Dentre as dificuldades da montagem de um centro de digitalização, certamente a mais trabalhosa e onerosa. O maior dos obstáculos está em conciliar a segurança da informação com a facilidade de acesso. No caso do IEB, as condições iniciais e o cenário tecnológico impuseram a realização destas cópias em mídias de CD/DVD, as quais hoje, em função do crescimento da produção e conseqüentemente do volume total

de arquivos digitais, demonstrou ser pouco eficiente. Optou-se, então, pela adoção de equipamentos de grandes capacidades de armazenamento digital em discos rígidos (*Storage*) e de unidades automatizadas de *back-up* em fitas LTO, procedimento que ainda não foi aplicado, uma vez que se encontra em fase de viabilização financeira, junto a instituições de fomento a pesquisa.

Disponibilização e extroversão de acervos digitais

A função primordial de uma digitalização é sua extroversão, seja através de sua reprodução em publicações, seja na possibilidade de acessos múltiplos e simultâneos, presenciais ou à distância, por meio de uma cópia digital. Não se trata de substituir o original por uma cópia digital, e sim de potencializar as possibilidades de acesso e, até mesmo, favorecer, em certa medida, a conservação do original pela redução do contato direto com ele.

Até o momento, falamos de como são geradas as matrizes digitais e das suas dificuldades de produção e manutenção, sempre tendo como foco a qualidade da imagem. No entanto, quando entramos no universo da extroversão, precisamos acrescentar mais uma variável a esta já complicada equação: a usabilidade desses arquivos. Isso porque nem sempre é possível distribuir a imagem no padrão da matriz digital para o público interessado, seja pela limitação do equipamento disponível ao interessado, seja pela limitação dos meios de distribuição. Surge, assim, a necessidade da geração de produtos específicos para atender as diferentes demandas. Por exemplo: se uma obra rara foi digitalizada em qualidade suficiente para uma eventual publicação *fac-similar*, para ser consultada em meio digital podemos reduzir sua resolução sem perda na qualidade de leitura, o que diminui a exigência do equipamento de consulta. No caso de uma disponibilização *on-line* via *internet*, essa redução é que garante a viabilidade de uso por um maior número de consulentes.

Existem diversos padrões que podem ser adotados para os produtos de extroversão. No IEB, optou-se pelo uso dos mais difundidos, a fim de potencializar a capacidade de difusão.

Certamente, por mais complexos e trabalhosos que sejam os processos de captura de imagens, de geração de matrizes digitais e de confecção dos produtos de extroversão, seu conjunto não representa a maior parte do tempo e do trabalho envolvidos na construção efetiva de um acervo digital funcional. A despeito de toda visibilidade conquistada por eles, sua utilidade/funcionalidade está intimamente ligada a dois outros processos essenciais: um, anterior a digitalização, é a seleção dos itens a ser digitalizados; o outro, o cadastramento deste acervo em uma base de dados passível de ser consultada para a localização das imagens.

Política de seleção do material e de sua extroversão

Nenhuma digitalização deve anteceder a definição de uma política clara de seleção e extroversão. É através das diretivas desta política que podemos definir o melhor equipamento e padrão de captura, os produtos gerados a partir das matrizes digitais, os *hardwares* para armazenamento e *back-up*, e disponibilização.

Definida a política de seleção, começa o trabalho de especialistas e pesquisadores na efetiva escolha do material a ser digitalizado. Novamente, não estamos diante de uma tarefa simples e, na maioria dos casos, a disponibilidade de equipamentos e/ou de pessoal mostra ser o grande limitador. Para melhor elucidar o problema, partiremos de um exemplo: a Biblioteca do IEB, com um montante aproximado de 180

mil volumes. Pela atual estrutura disponível, um acervo desta magnitude exige que o processo de digitalização seja fracionado, dividido em grupos que podem ser temáticos, cronológicos, genealógicos etc. É importante notar que a eficiência do resultado da extroversão está ligada à coesão dos grupos e ao seu percentual de digitalização. A experiência nos mostrou que a digitalização de forma pulverizada atrai a atenção de muitos, mas representa pouco resultado efetivo, ao passo que, quando a digitalização se concentra em um determinado agrupamento, se por um lado reduz o número de pesquisadores atendidos, por outro, transforma-se em uma ferramenta de pesquisa muito mais eficaz. Obviamente que, como tal, a eleição de grupos de obras a serem digitalizadas traz consigo os problemas inerentes a qualquer escolha.

Além das questões tratadas acima, devemos levar em conta alguns outros fatores importantes no processo de seleção das obras a serem digitalizadas:

1. Questões relacionadas a **direitos autorais**.
2. A **viabilidade técnica** de digitalização em função de características físicas dos documentos e equipamentos disponíveis.
3. As **condições de acesso**, priorizando aqueles que pela sua raridade têm seu acesso mais restrito dentro ou fora da instituição.
4. **Disponibilidade eletrônica**, evitando investir tempo para disponibilizar algo já disponível. A dificuldade é dada pela concomitância de projetos de digitalização e a quase ausência de comunicação entre eles.

Dessa maneira, podemos observar que o processo de seleção de obras para digitalização exige uma reavaliação constante, tanto ou mais que os processos de definição de padrões para captura.

Embora o exposto acima, nos dê uma idéia do panorama geral da digitalização no IEB, faz-se necessário ainda destacar alguns pontos em relação aos acervos do Arquivo e da Coleção de Artes Visuais. Ao tratarmos desta, um dos grandes desafios está ligado à própria natureza do acervo, o qual exige uma qualidade de imagem que busque manter ao máximo aspectos da cor e da textura do suporte, além de detalhes da técnica utilizada, em originais muitas vezes de grande formato e/ou tridimensionais.

Já no caso do Arquivo do IEB, a dificuldade está no gerenciamento das digitalizações. Se para biblioteca uma unidade (livro, periódico etc.) é composta por dezenas de imagens em suporte de papel, normalmente em tamanhos regulares, isto nem sempre se aplica ao acervo do Arquivo, onde um documento pode variar de um pequeno pedaço de papel a mapas com mais de 2m, de uma única folha a um caderno de manuscritos, passando por fotos, negativos, telas, xilogravuras etc., compondo um universo de mais de 450 mil documentos nos mais variados suportes e formatos, cada qual com suas necessidades específicas de informações para sua correta identificação. Visando a solucionar este problema, está em implantação no IEB um sistema de cadastramento eletrônico, onde os dados de cada documento são inseridos por especialistas antes da solicitação de sua digitalização, o que possibilitará no futuro sua extroversão.

Catálogo eletrônico IEB: histórico e descrição

Até o início de 2009, o setor de Arquivo do IEB dispunha de escassos instrumentos de pesquisa informatizados para a consulta dos documentos de seus acervos, frutos de projetos isolados de organização de fundos pessoais específicos. Entre eles, podemos citar o catálogo da correspondência do Fundo Mário de Andrade,

disponibilizado no site do IEB; o catálogo do Fundo Pierre Monbeig, acessível somente para consulta interna; os catálogos do Fundo Anita Malfatti e da correspondência do Fundo Aracy de Carvalho Guimarães Rosa, localizados na *intranet* da instituição, mas não utilizados. No mais, os instrumentos de pesquisa dos outros fundos organizados tomavam a forma de extensos catálogos impressos, de difícil manipulação e recuperação de informações.

Em vista desse quadro, e visando a um aprimoramento dos processos de organização e publicação dos seus acervos, foi iniciado um projeto de implementação e implantação de um catálogo eletrônico institucional conjunto para os setores do Arquivo e da Coleção de Artes Visuais do IEB que satisfizesse os seguintes propósitos:

1. Armazenasse informações descritivas das unidades constitutivas dos acervos existentes – documentos, obras de arte e objetos tridimensionais –, respeitando sua especificidade;
2. Armazenasse informações sobre as formas de agrupamento dessas unidades, de acordo com o processo de organização de cada acervo;
3. Disponibilizasse as informações armazenadas para o público consulente, por meio do site da instituição, permitindo a localização das unidades dos acervos através de uma ferramenta informatizada de busca.
4. Permitisse a adição, alteração ou exclusão de acervos ou unidades constitutivas, de acordo com as necessidades dos setores;
5. Exibisse para o consulente, quando disponível, imagens das unidades dos acervos.

Tecnologicamente, a construção do catálogo eletrônico exigiu a modelagem e implementação de um banco de dados relacional para o armazenamento das informações dos acervos, unidades constitutivas e seus agrupamentos, e de uma aplicação *web* que constituísse o *front-end* desse banco de dados, dividida em dois módulos: a) módulo de manutenção e gerenciamento de acervos, de uso interno; b) módulo de consulta, para ser disponibilizado para consulentes via Internet.

Um dos resultados da confecção do catálogo eletrônico foi alcançar uma padronização na forma como as unidades dos acervos são descritas. A partir disso, os responsáveis pela organização dos acervos podem inserir ou atualizar os dados dos documentos, obras ou objetos diretamente no repositório de dados.

Para os documentos do Arquivo, o catálogo sugere o preenchimento das seguintes informações descritivas básicas:

- Código de referência;
- Unidade de armazenamento;
- Espécie documental;
- Título (quando explícito);
- Suporte;
- Descrição;
- Local de produção;
- Data de produção;
- Autoria;
- Estado de conservação;
- Observações.

Para a Coleção de Artes Visuais, herdou-se o modelo descritivo já praticado no setor de maneira uniforme:

- Número de tombo;
- Título/Tema;
- Autor;
- Técnica;
- Descrição da obra;
- Data;
- etc.

Em especial, para os acervos do Arquivo, o catálogo eletrônico também permite o arranjo das unidades dos acervos de acordo com agrupamentos lógicos – séries, subséries, grupos, eventos, assuntos, etc. O Fundo Anita Malfatti, por exemplo, organiza-se em função de séries (Cadernos, Correspondência, Manuscritos, etc.) e subséries (da série Manuscritos: Biográficos/artísticos, Ensino, etc.).

Ao consulente está facultado o acesso às informações dos acervos por duas formas:

1. Visualização completa de todos os acervos do Arquivo ou da Coleção de Artes Visuais, agrupamentos lógicos e unidades constitutivas, exibidos em estrutura hierárquica (Setor → Acervo → (Agrupamentos lógicos) → Unidade do acervo), em navegação via hipertexto.
2. Busca de unidades de acervo específicas via critérios escolhidos pelo próprio consulente: palavras-chave, espécie documental, idioma, local de produção, data, etc.

Cada unidade constitutiva de acervo, portanto, é dotado de uma ficha descritiva única, acessível para consulta, e à qual podemos adicionar, potencialmente, inúmeras outras informações adicionais, conforme a necessidade. No estágio atual do catálogo eletrônico, a imagem do documento ou da obra de arte já pode ser anexada à sua ficha. Mas, eventualmente, com esforço relativamente reduzido, pode-se anexar à ficha outros tipos de arquivo: documentos relacionados, sons, vídeos, etc.

Adicionalmente ao trabalho de criação do módulo de manutenção e gerenciamento de acervos e do módulo de consulta do catálogo eletrônico, está em curso um contundente esforço de migração das informações dos acervos, originalmente armazenadas em bancos de dados avulsos e outros tipos de arquivos digitais. Como consequência desse trabalho de migração, já foi possível disponibilizar no catálogo eletrônico do IEB quase 80 mil registros.

O Catálogo Eletrônico IEB representa um grande avanço no modo como a instituição trata e disponibiliza seus acervos para consulta. À equipe técnica que organiza os acervos é oferecido um modelo de trabalho e uma ferramenta completa para descrição de acervos e suas unidades constitutivas. Ao consulente é exponenciada a facilidade de localização dos itens de seu interesse.

Acervos digitais: novos desafios

Nesse tipo de discussão, é fundamental estabelecer uma diferença básica: acervos digitalizados são aqueles que foram produzidos a partir das digitalizações de acervos físicos e correspondem a uma cópia digital daquele acervo original, enquanto

acervos digitais se referem ao universo do material produzido já em meio digital pelo próprio autor, desde documentos administrativos, registro fotográfico de eventos e/ou atividades, etc.

Além do aspecto histórico desta produção, atualmente a legislação exige a guarda destes arquivos:

São documentos de arquivo todos os registros de informação, em qualquer suporte, inclusive o magnético ou óptico, produzidos, recebidos ou acumulados.

(Decreto do Estado de São Paulo n. 48.897/2004, art. 3º)

As disposições deste decreto aplicam-se também aos documentos arquivísticos eletrônicos.

(Decreto do Estado de São Paulo n. 48.897/2004, art. 41)

Desta forma, fica colocada na ordem do dia a definição de estratégias de preservação dos acervos digitais, a fim controlar os riscos da degradação de suportes (mídias) e formatos, garantir também a manutenção do ambiente tecnológico mantendo a infra-estrutura necessária para acesso e utilização destes acervos.

É importante lembrar que não se trata de pura e simplesmente produzir um amontoado digital; é necessário que este seja tratado dentro das normas arquivísticas, assegurando a possibilidade de recuperação no futuro, não apenas das informações relativas ao conteúdo do documento como das informações sobre a geração do mesmo (metadados).

Sendo assim, estruturar um acervo digital exige planejamento arquivístico e infra-estrutura de informática complexos, na medida em que seu crescimento se dá de forma exponencial. Apenas para ilustrar, pensemos no caso deste texto que, apesar de apresentado ao leitor em sua forma impressa, possui um original digital em um programa de processamento de texto onde foi composto, uma versão digital em um *software* de diagramação onde adquiriu a sua forma final de impressão e, só por fim, sua versão impressa que representa, na verdade, uma cópia da versão original em meio digital. Lembremos ainda que esta é uma simplificação, pois voltando ao exemplo do próprio texto, ele foi composto da compilação de três outros arquivos de autores diferentes e, durante seu processo de preparação, foi distribuído, revisado, alterado, recompilado, gerando a cada passo deste processo, novos arquivos digitais igualmente importantes para o registro da memória. Portanto, esses arquivos digitais, como originais que são, devem ser tratados como documentos de memória, catalogados, arquivados e passar a integrar o acervo digital.

Conclusão

Vimos que os processos vinculados ao mundo digital têm, em sua própria essência, uma volatilidade que exige a constante reavaliação de estratégias e adequações tecnológicas para garantir sua sobrevivência e funcionalidade, tanto no que diz respeito a acervos digitais como digitalizados. Com o surgimento dos documentos digitais e digitalizados, alguns vislumbraram o fim dos acervos e de seus profissionais especializados, porém, ao longo deste texto, procuramos mostrar que a construção de um acervo digital só se faz com a participação ativa de arquivistas, bibliotecários, especialistas em imagens, profissionais de informática, e da instituição como um todo.

Desse modo, uma digitalização de acervos de magnitude e importância, como os que compõem o IEB, necessita de normas e padrões de qualidade a fim de garantir sua confiabilidade e sua extroversão, potencializando-os assim como fonte para pesquisas, o que temos empreendido no Instituto ao longo dos últimos anos.

Denis Machado Rossi
Frederico Antonio Camillo Camargo
Renato Cezar Muñoz