

# Resgate de conhecimento

Digitalização de acervos traz à tona raridades e documentos esquecidos e ajuda a aperfeiçoar o trabalho de pesquisadores

Fabrcio Marques



**A** multiplicação de projetos de digitalização de acervos de bibliotecas, arquivos e museus está modificando a forma como pesquisadores brasileiros trabalham. Nos últimos 15 anos, diversas instituições se empenharam em oferecer na internet documentos, fotografias e vídeos que antes só eram disponíveis em visitas agendadas. O resultado desse esforço é sensível. Em alguns casos, a facilidade de procurar e encontrar itens com ferramentas de busca ampliou acesso a informações difíceis de garimpar manualmente, potencializando a qualidade da pesquisa. Em outros exemplos, permitiu ao menos conhecer remotamente a amplitude de determinado acervo para organizar uma consulta presencial mais rápida e eficiente. “Estudantes e pesquisadores estão sendo formados nesse novo contexto. É um caminho sem retorno”, diz o historiador Pedro Puntoni, professor da Universidade de São Paulo (USP) e pesquisador do Núcleo de Cultura Digital do Centro Brasileiro de Análise e Planejamento (Cebap). “A biblioteca e o arquivo em papel sempre vão ter importância, mas perdem espaço para a internet diante da facilidade de ter acesso a documentos, imagens e livros, assim como teses e revistas digitais disponíveis *on-line*”, afirma.

De um computador pessoal em sua casa, o professor e pesquisador Wilton José Marques encontrou em meados de março um poema esquecido de um dos principais nomes da literatura brasileira, o escritor Machado de Assis (1839-1908), conhecido do leitor sobretudo pelos contos e romances. Não se trata de um poema qualquer, mas o primeiro, publicado no jornal *Correio Mercantil*, do Rio, em 9 de setembro de 1856, intitulado “O grito do Ipiranga”. Especialistas da obra de Machado consideravam que sua produção iniciara em 1858 com o poema “Esperança”, quando o autor, aos 19 anos, começou a trabalhar como revisor naquele jornal. O achado de Marques foi possível graças à digitalização do *Correio Mercantil* pela Biblioteca Nacional, que criou em 2009 sua hemeroteca digital ([bndigital.bn.br/hemeroteca-digital/](http://bndigital.bn.br/hemeroteca-digital/)). Marques, professor do Departamento de Letras da Universidade Federal de São Carlos (UFSCar), diz que encontrou a poesia graças a “um pouco de faro e um pouco de sorte”. Procurava os primeiros poemas de Machado para uma pesquisa sobre influências românticas na obra do autor. “Fui conferir as fontes, olhando poema por poema das coletâneas, a partir de 1858. Por curiosidade, resolvi fazer a busca em anos anteriores e ‘O grito do Ipiranga’, apareceu”, diz. Marques decidiu interromper sua pesquisa e se dedica a um artigo sobre a poesia. “É um poema longo, uma glorificação do grito da independência e de dom Pedro I. Uma das características da literatura de Machado é a intertextualidade: ele dialoga com outras obras e referências históricas. Nesse primeiro poema, compara o tempo todo a Independência com a república romana”, diz Marques. Outro interesse do pesquisador é lançar luzes sobre a juven-

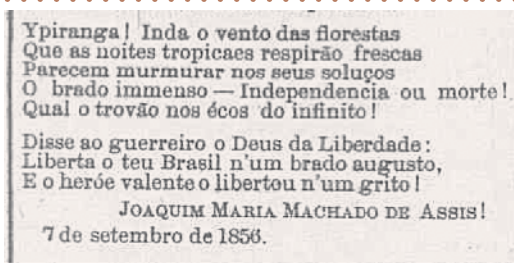
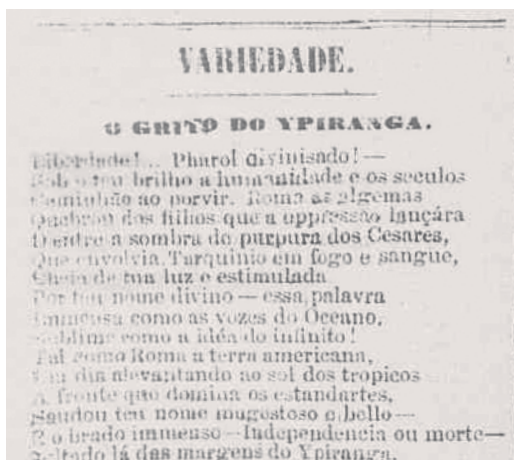
Digitalização de livros da Biblioteca Brasileira da USP: 4,8 mil volumes da coleção do bibliófilo José Mindlin já estão disponíveis *on-line*

tude de Machado. “Imagine como numa sociedade escravocrata um jovem negro de 17 anos, com educação formal que a gente não sabe muito bem como foi obtida, conseguiu se inserir no universo intelectual do Rio de Janeiro e colaborar num jornal importante da época.”

A Biblioteca Nacional tem um dos mais longevos programas de digitalização no país. Começou em 2006 e, hoje, oferece 900 mil documentos *on-line*, que rendem 400 mil consultas virtuais por mês. Há coleções de fotos, mapas e músicas. No mês passado, foi lançado o portal Brasileira Fotográfica ([brasilianafotografica.bn.br](http://brasilianafotografica.bn.br)) com mais de 2 mil fotos históricas de coleções da própria biblioteca e do Instituto Moreira Salles. A maior parte do acervo digital é composta por jornais brasileiros. São 5 mil títulos, digitalizados com recursos da Financiadora de Estudos e Projetos (Finep). “Temos a prerrogativa do depósito legal, que é a recepção de um exemplar de todas as publicações produzidas em território nacional. Por isso, nossa coleção é a mais abrangente do país”, diz Angela Bettencourt, coordenadora da Biblioteca Nacional Digital. A decisão de oferecer os jornais se deveu também a uma questão prática: o acervo era dos mais consultados da instituição.

A consulta aos jornais é simples – uma busca por palavra é suficiente para encontrar o que há sobre aquela referência. Para pesquisadores, a utilidade vai muito além da chance hipotética de encontrar um poema esquecido. O sociólogo Benno Warken Alves, de 25 anos, concluiu em 2014 sua pesquisa de mestrado, feita na USP com bolsa da FAPESP, sobre a trajetória de um empresário negro em Curitiba no século XX, Sydnei Lima Santos (1925-2001). Alves conseguiu referências sobre a trajetória de antepassados do empresário consultando na hemeroteca jornais de Sergipe, Rio de Janeiro e Paraná. “Se tivesse de procurar em microfilmes, não teria dado tempo ou talvez nem encontrasse.”

**S**e uma hemeroteca digital é fonte suficiente para levantar e conferir dados, a pesquisa de documentos históricos raramente dispensa a análise do arquivo físico, embora a digitalização consiga acelerar esse trabalho. A arquiteta Suely Figueirêdo Puppi tem utilizado o acervo digital do Instituto Lina Bo e Pietro Maria Bardi para sua tese de doutorado, na Universidade Federal do Rio Grande do Sul, sobre as técnicas de restauro e de projeto de Lina Bo Bardi (1914-1992), a arquiteta ítalo-brasileira conhecida por projetar o Museu de Arte de São Paulo (Masp). “Os projetos e desenhos de Lina disponíveis no *site* não têm a definição necessária para usar num livro, mas para uma tese são suficientes”, afirma Suely. Desde que começou o doutorado, em 2012, ela visitou algumas vezes a sede do instituto, em São Paulo,



“O grito do Ipiranga”, poema esquecido de Machado de Assis publicado no jornal *Correio Mercantil*, encontrado graças à hemeroteca digital da Biblioteca Nacional

para analisar documentos. “Faço uma pré-seleção *on-line* e já chego sabendo o que preciso”, afirma.

A catalogação e a digitalização do acervo também foram úteis na comemoração do centenário de Lina. “Os desenhos foram de grande valia para os curadores das mostras sobre sua obra realizadas em Zurique e em Munique, além de duas exposições em Nova York em que Lina teve representação. A primeira avaliação do acervo foi feita a distância”, diz Renato Anelli, professor do Instituto de Arquitetura e Urbanismo da USP em São Carlos, pesquisador responsável pelo projeto de digitalização. Os projetos e desenhos de Lina são o ponto forte do acervo digital. “Ela coloria e criava cenas e perspectivas nos projetos arquitetônicos. São desenhos fabulosos”, afirma Anelli. É possível consultar no banco de dados fichas sobre fotos e documentos escritos da arquiteta. O instituto começou a organizar o arquivo de Pietro Maria Bardi (1900-1999), criador do Masp. Bardi e Lina foram casados por 45 anos.

As principais limitações dos acervos digitais são a sua impossibilidade de oferecer a experiência sensorial de ver ou manusear um documento histórico e também a dificuldade de fornecer todas as informações necessárias para contextualizar as circunstâncias em que cada documento foi produzido e armazenado. “Documentos de arquivos públicos têm uma peculiaridade. Eles fazem sentido dentro do contexto em que foram produzidos. Longe desse contexto, não são compreendidos na sua integridade”, diz Marcelo Chaves, diretor do Centro de Difusão e Apoio à Pesquisa do Arquivo Público do Estado de São Paulo. A possibilidade

de extrair informação mais íntegra de um documento de arquivo, ele explica, está em conhecer a sua trajetória, a finalidade com que foi produzido e por onde circulou. “Sem isso, o documento perde parte de seu potencial informativo”, afirma. O trabalho do historiador Bruno de Andréa Roma, que atuou como estagiário do arquivo por dois anos, mostra esse desafio. Durante a graduação na USP, ele produziu, sob orientação do professor Carlos Bacellar, ex-coordenador do arquivo, um guia de fontes de tudo quanto existe sobre a Universidade

## Lacunas na catalogação de documentos digitais podem enviesar resultados de pesquisa

de São Paulo naquele grande acervo. “Guias e inventários são instrumentos fundamentais para se ter consciência do que existe num arquivo sobre determinado acervo, além de conhecer a trajetória dos documentos”, afirma Bruno, que agora faz mestrado sobre a fotografia no ambiente do arquivo. “A fotografia não tem um lugar tão bem situado quanto os demais documentos. No Arquivo do Estado, os negativos de coleções como a do jornal *Última Hora* estão num lugar, mas contatos e ampliações fotográficas, que trazem informações sobre o seu uso, estão em outro e não conversam entre si.”

Digitalização de mapas históricos no Arquivo Público do Estado de São Paulo: preocupação em oferecer metadados dos documentos



Marcelo Chaves ressalta que a digitalização foi fundamental para democratizar a informação e sem ela não seria possível constatar, hoje, que franquear documentos na internet não é suficiente para torná-los realmente acessíveis. “As políticas de digitalização são irreversíveis, mas precisam ter critérios”, diz. Nos últimos tempos, Chaves foi incumbido pelo arquivo de orientar prefeituras e câmaras municipais paulistas e encontrou situações desastrosas. “Muitos prefeitos são convencidos por empresas especializadas de que a digitalização é um benefício em si e acabam transformando arquivos físicos desorganizados em acervos digitais inacessíveis, gastando muito dinheiro”, afirma.

Nos últimos anos, o Arquivo Público do Estado de São Paulo dedicou-se a disponibilizar seu acervo

na internet, tendo como carro-chefe o arquivo do Departamento Estadual de Ordem Política e Social de São Paulo (Deops), principal órgão da polícia política paulista, extinto em 1983. Há dois anos, lançou o portal Memória e Resistência, que permite a consulta pela internet de mais de 314 mil fichas e 12,8 mil prontuários – num total de 1 milhão de imagens – produzidos por órgãos de vigilância política entre 1924 e 1999. Os técnicos do arquivo avaliam que parte desse trabalho precisa ser corrigido e até refeito. Ocorre que, da forma massiva como foi feita a digitalização, informações relevantes sobre os documentos oferecidos não estão disponíveis para os usuários. Isso pode não ser um problema para quem apenas quer consultar o que consta em seu nome ou de algum familiar dentro do arquivo. Mas, no caso de pesquisadores, as lacunas são suficientes para enviesar resultados, observa o historiador Marcelo Quintanilha Martins, diretor do Centro de Acervo Permanente do arquivo. O acervo do Deops é formado por três arquivos distintos, pertencentes às suas delegacias especializadas: Ordem Social, Ordem Política e Serviço Secreto. Frequentemente, documentos estão interrelacionados. Mas isso nem sempre é perceptível na pesquisa *on-line*. “O Deops dialogava, por exemplo, com o FBI norte-americano, do qual recebia memorandos relacionados a ordens de prisão. Mas esse contexto não está disponível nos documentos digitalizados.” Nas cerca de 50 mil imagens digitalizadas atualmente pela equipe do arquivo a cada mês, os chamados metadados, informações que descrevem o que pode ser encontrado nos documentos, são catalogados utilizando um *software* de acesso aberto, o ICA-AtoM, que permite armazenar um conjunto abrangente de dados que ajudam a prevenir a perda do contexto.

A ideia é revisitar documentos já digitalizados, enquadrando-os nesse padrão.

A padronização dos dados não é uma discussão restrita a arquivos e bibliotecas. Giselle Beiguelman, professora da Faculdade de Arquitetura e Urbanismo da USP e organizadora do livro *Futuros possíveis – Arte, museus e arquivos digitais*, editado com apoio da FAPESP, chama a atenção para a figura do “coleccionador *on-line*”, aquele indivíduo que oferece na internet documentos, imagens ou vídeos de valor histórico. “Com generosidade e investimento de tempo, eles disponibilizam fotos de família, memórias pessoais ou vídeos no YouTube com programas de TV da década de 1970 e filmes históricos desde o início dos anos 1920, que podem ter utilidade para pesquisadores, mas precisariam de padronização dos seus processos de catalogação e seus metadados”, diz Giselle. Ela menciona os problemas relacionados à “corporativização da memória”, que é o arquivamento de imagens e documentos em plataformas privadas, como Flickr e YouTube. “Cada uma delas organiza documentos de uma forma e há o risco de que sejam tirados do ar a qualquer momento, de acordo com o interesse das empresas. Não há uma discussão sobre sistemas de metadados padronizados para todos, inclusive as plataformas privadas”, afirma. A ideia de criar padrões, ela diz, é factível. “A internet é o melhor exemplo. Funciona a partir de protocolos e uso regrado de símbolos comuns, que todos seguem, como a arroba no endereço de *e-mail*.” Um dos artigos do livro, assinado por Monika Fleischmann e Wolfgang Strauss, do Mars-Exploratory Media Lab, da Alemanha, cita o conceito de mapa

semântico de conhecimento como possibilidade para registrar e visualizar um conjunto de informações acerca de um documento. “As entradas individuais dos arquivos são localizadas de modo relacional e os parentescos semânticos são visualizados”, dizem os autores.

**A** digitalização de acervos tem avançado mundo afora desde o início dos anos 2000. “Nos Estados Unidos, as universidades comandaram o processo de digitalização de bibliotecas”, diz Pedro Puntoni, que dirigiu a Biblioteca Brasileira Guita e José Mindlin da USP entre 2007 e 2014 e desenvolveu a Biblioteca Brasileira Digital. “Já a União Europeia financiou um consórcio para fortalecer as bibliotecas virtuais, capitaneado pela Gallica, da biblioteca nacional da França.” O Brasil, observa Puntoni, teve certo pioneirismo na década passada ao catalogar, microfilmar e digitalizar cerca de 3 milhões de páginas sobre os primeiros 300 anos de história do Brasil pertencentes ao Arquivo Histórico Ultramarino, de Portugal. O trabalho foi coordenado por Esther Bertoletti, da Biblioteca Nacional. Mas a digitalização de acervos no país foi prejudicada pela dificuldade de transformar projetos em programas permanentes. “Em muitos casos, equipes são contratadas por tempo determinado e se desarticulam quando há descontinuidade no financiamento ou no comando da instituição”, diz Puntoni.

Em 2009, a FAPESP lançou uma chamada de propostas do Programa de Apoio à Infraestrutura de Pesquisa, na modalidade Museus e Centros Depositários de Informações e Documentos e de Coleções Biológicas, que selecionou 20 projetos,



## A redenção vai à internet

Coleção de jornal abolicionista é recuperada

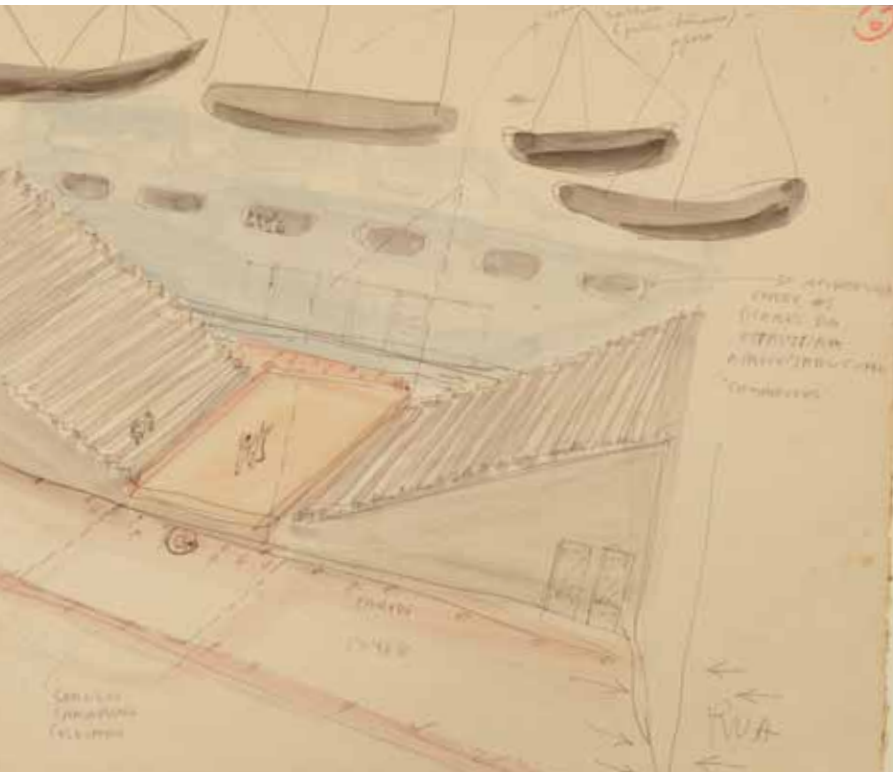
O Arquivo Público do Estado de São Paulo vai lançar no segundo semestre um *site* com 132 edições digitalizadas do jornal abolicionista *A redenção*, que circulou em São Paulo de 2 de janeiro de 1887 até a promulgação da Lei Áurea, em 13 de maio de 1888 e era porta-voz do movimento abolicionista dos “caifazes”, grupo que resgatava escravos, levando-os a locais seguros. “Era um jornal radical, com ataques a fazendeiros, políticos e a outros jornais, mesmo aqueles abolicionistas”, diz Marcelo Quintanilha Martins, diretor do Centro de Acervo Permanente do Arquivo.

A coleção digitalizada veio do Instituto Histórico e Geográfico de São Paulo e boa parte dos exemplares estava em pedaços. Os restauradores montaram os fragmentos do jornal com o auxílio de pinças, consultando cópias de microfilme dos exemplares existentes na Biblioteca Lamont, da Universidade Harvard, nos Estados Unidos. Nesse processo, foram encontrados sete exemplares que se imaginavam perdidos. A coleção foi incluída, no final de 2014, no Registro Nacional do Brasil do Programa Memória do Mundo,



da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco). O programa reconhece como patrimônio da humanidade documentos, arquivos e bibliotecas de importância internacional, regional e nacional. O objetivo é ampliar a difusão dos acervos.

Edição histórica da abolição da escravatura do jornal: coleção digitalizada será oferecida no segundo semestre



Teatro do Sesc Pompeia em croqui de Lina Bo Bardi: *site* reúne desenhos da arquiteta modernista

muitos dos quais voltados para a organização, digitalização e oferta de documentos *on-line*. Foram contempladas instituições como o Instituto Lina Bo e Pietro Maria Bardi, o Arquivo Público do Estado de São Paulo, o Museu Lasar Segall e o Instituto de Estudos Brasileiros da USP, entre outros. Órgãos como o Banco Nacional de Desenvolvimento Econômico e Social (BNDES), a Petrobras e a Finep também se destacam na oferta de recursos para projetos de digitalização. Um grupo de instituições comprometidas com políticas de digitalização no Brasil reuniu-se na Rede Memorial para compartilhar experiências e ajudar a organizar acervos, além de desenvolver uma ferramenta de busca que reúna em um só lugar acervos digitalizados de todas as bibliotecas nacionais.

“Houve na década passada uma boa oferta de recursos para projetos de digitalização no país, ao mesmo tempo que os equipamentos se tornaram mais baratos”, diz Millard Schisler, pesquisador e consultor na área de digitalização e preservação digital. Mas a corrida para oferecer acervos *on-line*, afirma ele, gerou distorções. “Era comum encontrar instituições que investiam muito em digitalização, mas pouco em conservação dos documentos originais, que ficavam acondicionados de forma precária”, diz Millard. Hoje, buscam-se estratégias sustentáveis. “Em vez de querer digitalizar tudo e em alta definição, é mais razoável oferecer as partes mais procuradas de um acervo, reservando recursos para a preservação dos originais. A digitalização deve ser vista como um complemento, não como a estratégia principal.”

O exemplo do Museu Imperial, de Petrópolis, resume os desafios de digitalizar acervos. A empreitada começou em 2009. Hoje, 8 mil itens estão disponíveis no *site* da instituição, entre documentos, livros e imagens de itens museológicos, o equivalente a 3% do acervo, e geram 2 mil acessos por mês. “A intenção é digitalizar a totalidade do acervo”, diz o historiador Jean Bastardis, coordenador da equipe contratada pela sociedade dos amigos do museu e incumbida de pesquisar documentos, descrevê-los, organizar a base de dados e disponibilizar o material no *site*. A prioridade é digitalizar conjuntos completos de coleções para facilitar o trabalho dos pesquisadores. Apesar dos seis anos de experiência, manter o processo de digitalização tem sido uma missão acidentada. A equipe é contratada por projeto. A cada ano repete-se o desafio de obter recursos de empresas ou de editais públicos.

Entre os itens mais procurados no *site*, destaca-se um livro de receitas do século XIX, *O cozinheiro imperial*, que descreve pratos da época e cardápios de banquetes. Pesquisadores têm mais interesse pelo arquivo da Casa Imperial, com documentos produzidos em Portugal que remontam ao século XIII. A experiência do museu mostra que o trabalho de divulgação é essencial. “Quando sai uma reportagem sobre algum item, a procura por ele tem picos”, diz Bastardis.

A Biblioteca do Congresso dos Estados Unidos, com mais de 20 milhões de objetos digitais disponíveis *on-line*, ao constatar que a oferta na internet não é suficiente para disseminar o conhecimento, criou um programa pelo qual seus técnicos vão a cidades ensinar professores e bibliotecários locais a explorar seu acervo. Sem boa divulgação, acervos *on-line* não alcançam todo o público potencial. A digitalização de 100 horas de dramaturgia da extinta TV Tupi depositadas na Cinemateca Brasileira abriu caminho para estudos sobre novelas que antes eram inviáveis. Mas o material, disponível no *site* [www.bcc.org.br/tupi/telenovelas](http://www.bcc.org.br/tupi/telenovelas), ainda é pouco explorado, diz Esther Hamburger, professora da Escola de Comunicações e Artes da USP. Ela foi a pesquisadora responsável pelo projeto que catalogou e digitalizou as 100 horas de gravação, e vem realizando estudos com o material.

Segundo Esther, o arquivo da Tupi está num lugar ainda pouco visível. “É preciso ter estrutura de rede que possa receber muitos acessos. O banco de dados é sólido nesse sentido, mas o projeto precisa ser completado”, diz. O conteúdo da televisão deveria estar mais disponível *on-line*, afirma a pesquisadora. “Países como França e Suécia disponibilizam o conteúdo que foi ao ar na TV através de suas bibliotecas nacionais”, afirma. Ela pondera que ainda não existe no Brasil tradição em pesquisa histórica utilizando material que a televisão levou ao ar.

**A**s instituições também aprenderam que um projeto de digitalização, uma vez concluído, não chegou realmente ao fim. É preciso investir em manutenção e atualização tecnológica. “No caso de filmes, manter um arquivo digital pode custar até 10 vezes mais do que manter um arquivo físico”, diz o consultor Millard Schisler, citando o estudo *O dilema digital*, feito pelo Science and Technology Council da Academy of Motion Picture Arts and Sciences, de Hollywood, que comparou a manutenção de filmes em película e digitais. É preciso criar e manter bancos de dados, atualizar *softwares* e monitorar o estado dos documentos. A cada vez que um acervo migra para outro servidor ou formatos digitais são atualizados, estima-se que até 10% dos documentos se percam – e cabe a um gestor recuperá-los a partir dos originais.

Inaugurada na USP em março de 2013, a Biblioteca Brasileira Guita e José Mindlin incorporou no ano passado os volumes digitalizados pelo Projeto Brasileira USP e criou seu acervo digital. Já foram digitalizados 4,8 mil dos 32 mil volumes do acervo composto por livros, manuscritos e periódicos raros garimpados ao longo de 80 anos pelo empresário José Mindlin (1914-2010). O trabalho de digitalização na biblioteca passa agora por uma atualização. Ocorre que boa parte dos livros disponíveis *on-line* foi digitalizada e submetida a um processo pelo qual a imagem colorida é convertida para apenas dois tons. O processo criou algo parecido como uma cópia xerox do original, em que o texto aparece sobre um fundo branco. A decisão foi tomada na época pelo Projeto Brasileira USP para gerar imagens mais limpas e arquivos leves, facilitando o acesso para usuários com internet mais lenta e permitindo a economia de tinta em caso de impressão dos arquivos pelos usuários. Hoje, explica Jony Fávoro, especialista em digitalização da biblioteca, o método é diferente. Busca-se digitalizar as obras preservando características como o amarelado das páginas, as marcas de uso e a reprodução fiel da capa. “Para um usuário que busca apenas o texto da obra, pode não fazer diferença”, afirma Sandra Guardini Vasconcelos, diretora da biblioteca e professora do Departamento de Letras Modernas da Faculdade de Filosofia, Letras e Ciências Humanas da USP. “Mas, no caso de um historiador, uma reprodução fidedigna traz elementos importantes sobre a história editorial daquela obra.” A biblioteca planeja redigitalizar as primeiras obras, embora a prioridade seja oferecer as ainda não disponíveis. “Essa preocupação com o livro no seu suporte material, isto é, como objeto físico, é seguida por grandes instituições, como a British Library e a Bibliothèque Nationale de France”, diz Sandra. ■



1 Bandeira do Brasil no período monárquico, do acervo digital do Museu Imperial



2 Fotografia de veste do imperador Pedro II disponível *on-line*

### Projetos

1. Acervo quadruplex da extinta TV Tupi (nº 2009/54923-7); **Modalidade** Auxílio à Pesquisa – Programa Infraestrutura; **Pesquisadora responsável** Esther Império Hamburger (USP); **Investimento** R\$ 446.934,77 (FAPESP).
2. Acervo do Instituto Lina Bo e Pietro Maria Bardi: catalogação, digitalização e montagem de banco de dados *on-line* (nº 2009/54901-3); **Modalidade** Auxílio à Pesquisa – Programa Infraestrutura; **Pesquisador responsável** Renato Luiz Sobral Anelli (USP); **Investimento** R\$ 253.269,46 (FAPESP).
3. Preservação e difusão da memória pública: modernização e ampliação dos laboratórios do Arquivo Público do Estado de São Paulo (nº 2009/54965-1); **Modalidade** Auxílio à Pesquisa – Programa Infraestrutura; **Pesquisador responsável** Carlos de Almeida Prado Bacellar (USP); **Investimento** R\$ 1.692.982,33 (FAPESP).
4. Por uma Biblioteca Brasileira digital (nº 2007/59783-3); **Modalidade** Auxílio à Pesquisa – Regular; **Pesquisador responsável** Pedro Luis Puntoni (USP); **Investimento** R\$ 663.514,35 (FAPESP).