

Universidade de São Paulo
Pró-Reitoria de Cultura e Extensão Universitária
Projeto Aprender com Cultura e Extensão

Projeto

O processamento digital de imagens e a descrição de documentos iconográficos na *Brasiliana Digital*

Maria Clara Paixão de Sousa
Departamento de Letras Clássicas e Vernáculas
Faculdade de Filosofia, Letras e Ciências Humanas
Universidade de São Paulo

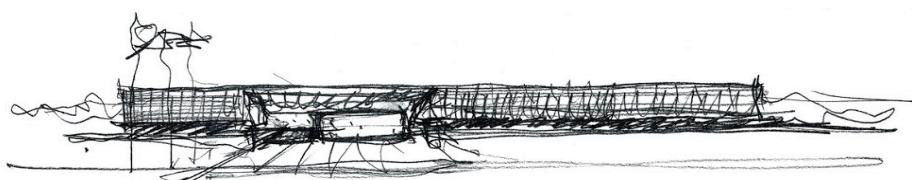
Daniela Pires
Biblioteca Brasiliana Guita e José Mindlin
Universidade de São Paulo

Conteúdo

1. Introdução	3
2. Objetivos.....	4
3. Justificativas.....	4
4. Metodologia e Plano de Trabalho	5
4.1 Metodologia.....	5
4.2 Plano de Trabalho.....	8
4.2.1 Capacitação técnica e teórica	8
4.2.2 Trabalho de descrição dos documentos e produção de metadados	8
4.2.3 Avaliação e apresentação de resultados.....	8
5. Cronograma.....	9
6. Referências Bibliográficas.....	9

1. Introdução

Esta proposta de pesquisa insere-se no contexto maior dos projetos *Brasiliana USP*¹ e *Brasiliana Digital*. O Projeto *Brasiliana* é uma iniciativa da Reitoria da Universidade de São Paulo, com a missão de custodiar e desenvolver a *Biblioteca Brasiliana*, reunindo cerca de 500.000 volumes de inestimável valor histórico, fruto da união entre os acervos do *Instituto de Estudos Brasileiros*, órgão com a tradição de mais de 40 anos dedicados à curadoria de material histórico na USP, e da *Biblioteca Brasiliana Guita e José Mindlin*, fundada em janeiro de 2005² como abrigo da coleção reunida pelo bibliófilo José Mindlin e generosamente doado à USP em maio de 2006. Os acervos serão transferidos em conjunto para o edifício especialmente construído para este fim no coração da Universidade, previsto para ser terminado em 2012.



Maquete do Novo Edifício da Biblioteca Brasiliana USP

A esta nova condição física, favorecedora da consulta pública, soma-se a *Biblioteca Brasiliana Digital* (BBD), que permite o alcance ao conteúdo dos dois acervos para fins de pesquisa geral e acadêmica, sob os moldes propostos no projeto “*Por uma Biblioteca Brasiliana Digital*” (Puntoni, 2007)³. A fase piloto de implantação da Brasiliana Digital contou com o apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp), por meio de um auxílio à pesquisa (processo 07/59783-3), e de diversas outras instituições, grupos de pesquisa e pesquisadores.



O Laboratório da Brasiliana Digital

No início de 2009, foi inaugurado, junto ao canteiro de obras do edifício em construção, o *Laboratório Brasiliana*, que atualmente serve de abrigo para uma equipe de pesquisadores e bolsistas de diferentes cursos de graduação e pós-graduação da USP (História, Letras, Engenharias, Matemática e Computação), empenhados, sob a coordenação dos pesquisadores docentes, em enfrentar os diversos desafios envolvidos na produção de um acervo digital a partir de obras antigas e raras.

¹ cf. <<http://www.brasiliana.usp.br>>

² cf. Resolução da Reitoria da Universidade de São Paulo Nº 5172, 23.12.2004. D.O.E, 24.12.2004; <http://leginf.uspnet.usp.br/resol/r5172m.htm>

³ Projeto financiado pela *Fundação de Amparo à Pesquisa do Estado de São Paulo*, 07/59783R (Auxílio Regular à Pesquisa), do qual a proponente do presente Projeto participa como pesquisadora colaboradora.

Na proposta aqui apresentada, buscam-se alunos de graduação interessados em trabalhar na área tratamento de imagens digitais e na descrição bibliográfica de documentos iconográficos antigos e raros em colaboração direta com a equipe de biblioteconomia da *Brasiliana*. O estágio possibilitará aos alunos um contato com a experiência prática do trabalho de tratamento de imagens utilizando vários tipos de *softwares*, o aprendizado sobre o processo de catalogação e descrição de documentos digitais e lhes oferecerá uma oportunidade de colaborar para um grande projeto de pesquisa coletivo.

2. Objetivos

O projeto tem como meta específica o processamento digital das imagens e a produção de descrições dos documentos iconográficos das obras do acervo *Brasiliana*, que por possuírem características particulares demandam um tratamento diferenciado. O objetivo geral, a partir disso, é colaborar para a uniformização do catálogo de iconografia do acervo *Brasiliana-USP* e para a ampliação do acesso a esse catálogo por parte de mecanismos de busca. As atividades no estágio deverão incluir:

- O processamento digital das imagens e obras digitalizadas utilizando ferramentas e softwares específicos;
- A descrição dos documentos iconográficos que compõem o acervo da coleção *Viajantes da Brasiliana Digital*;
- A aplicação de metodologia de descrição bibliográfica de imagens digitais que norteará a descrição de documentos iconográficos.

3. Justificativas

As bibliotecas digitais se oferecem como um instrumento de multiplicação, de universalização de acesso, e de democratização dos meios. Sendo assim, devemos procurar um modelo de Biblioteca Digital baseado em padrões atualizados e compartilhados de produção de objetos (imagens e textos) e de descrição (metadados), com condições de ofertar títulos com rapidez. Padrões necessários para uma biblioteca que se quer aberta e ajustada a sua finalidade, ou seja, atender aos seus usuários, oferecendo-lhes um instrumento efetivo de **pesquisa, de estudo e de formação**.

Com o seu expressivo conjunto de livros e manuscritos (cerca de 40.000 volumes), a *Brasiliana* de Mindlin é considerada a mais importante coleção do gênero formada por um particular, comportando obras de literatura brasileira e portuguesa, relatos de viajantes, manuscritos históricos e literários (originais e provas tipográficas), periódicos, livros científicos e didáticos (cf. Mindlin, 2005). Entre as inúmeras preciosidades, destaca-se uma das mais completas coleções de obras do século XVI ao XIX escritas por viajantes pelo interior do Brasil, constituída em sua maioria por textos escritos em outros idiomas, onde os mais significativos são: alemão, latim e francês. Estes livros trazem em seu conteúdo uma infinidade de documentos iconográficos, que retratam o Brasil em diferentes épocas.

Um dos objetivos deste Projeto é oferecer aos usuários a oferta desse material iconográfico, por meio de um banco de dados de imagens. Para que a *Biblioteca Brasiliana Digital* atinja seu objetivo (disponibilização da informação) um dos requisitos necessários está relacionado à descrição bibliográfica minuciosa destas informações iconográficas, esta descrição está relacionada à produção de metadados, que devidamente organizados e estruturados garantirão que a informação seja descoberta, localizada e possa ser utilizada. O desafio hoje está na descrição da informação que se encontra em diversos formatos digitais e que precisam atender a diferentes públicos e usos. Para auxiliar no processo de recuperação destes documentos, é necessária a descrição de algumas informações importantes como: determinar o

ilustrador, gravador ou desenhista da imagem ou conjunto de imagens, título da imagem, imprensa, legendas, relação entre o livro e seu conjunto de imagens e vice-versa além de outras informações complementares que permitirão ao público em geral o acesso a estes conteúdos de forma plena e eficiente, justificando-se então a extensão da *Brasiliiana Digital*, cujo público hoje está na média de 1.000 (hum mil) acessos/dia.

Aos bolsistas participantes do projeto aqui proposto, o contato direto com esta coleção de documentos permitirá um desenvolvimento acadêmico em duas frentes: de um lado, o conhecimento/aprimoramento do seu conhecimento sobre processamento de imagens digitais por meio de *softwares* e ferramentas específica; de outro, o aprendizado de diferentes técnicas descrição e organização de repositórios e objetos digitais. O trabalho junto a um grande projeto de pesquisa, como é o *Brasiliiana USP*, oferecerá ainda aos alunos a oportunidade de participar de perto das atividades acadêmicas ligadas ao laboratório (como seminários e oficinas), favorecendo assim uma formação interdisciplinar – reforçada pela convivência e interação com alunos e pesquisadores de diferentes unidades da Universidade. Nesse sentido, ressalte-se que o trabalho em conjunto com a bibliotecária responsável da *Brasiliiana Digital* permitirá, aos alunos, um contato com uma área do conhecimento interdisciplinar, permitindo assim uma aproximação inicial que poderá ser útil na ampliação de seus horizontes profissionais. Destaca-se, por fim, o aspecto da função social do trabalho destes alunos, que colaborarão na facilitação do acesso aos conteúdos de uma biblioteca pública, cumprindo assim uma das premissas da experiência universitária na USP: a integração entre ensino, pesquisa e extensão.

4. Metodologia e Plano de Trabalho

4.1 Metodologia

Uma biblioteca digital pode ser definida como um conjunto de documentos sistematizados para acesso mediado no meio digital, preparados para reconhecimento e extração automática de informação. O seguinte diagrama ilustra os processos envolvidos na “digitalização” – isto é: na captação de informação não-digital e geração de informação digital:

Diagrama: Etapas na formação de um acervo digital

1. Seleção no Acervo Físico



2. "Digitalização"

(i) Geração de Arquivos de Imagens



História und Beschreibung eynner Landschaft der Wölken / Niederlanden / Graueischen / Oostfriesischen / Leuven / in der Nieuwemel / Aperten / gelegen / vor und nach Graft / gehört im Land zu Delft / und ist / von vñ diese / nicht vergangen / ist / Da se Hans Staden von Hornsberg / auf Delft durch sein eyne erfahrung erkannt / und weg durch den ruck an tag gibt /

Dedictie den Durchleuchtungen Hodgeboonen beren / D'Philippen Landgraaf in Delft / Graaf van Ledenburg / D'Uit / Steeghman van 's Hulde / seinn G. D.

Mit einer vorsichtige D. J. d. Brandenburgian Schman / Oordt / Prograaf / Medel / s' Steeghman.

Gebr. der Delftina volget nach den Delftina.

Jpeg, png, tiff

(ii) Geração de Documentos de Acesso (textos)

- (a) Pós-processamento de imagens
- (b) Geração de arquivos de texto
- (c) Tratamento para extração de Conteúdo (*Catalogação, Revisão, etc*)

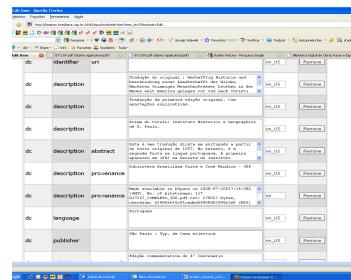
Scanner

Txt, doc, xml, html., pdf...

3. Preparação dos metadados dos livros

Extração de Informação (Conteúdos)

Inserção das informações de interesse no banco de dados, por exemplo: tradução do título, se possui imagens, quando livro for ilustrado, inserção das informações também no banco de imagens, onde cada imagem será catalogada separadamente.

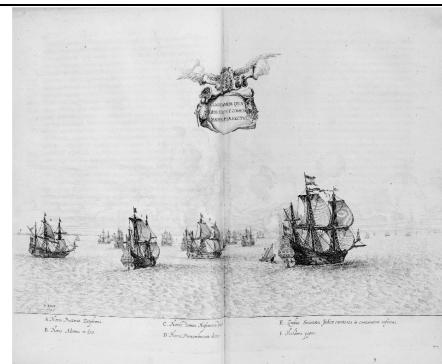


Portal Web com documentos PDF, Html...

4. Extração de metadados das imagens

Extração de Informação (Conteúdos)

Extração das informações de interesse no banco de dados, por exemplo: tradução do título da imagem, tradução das legendas, para servirem como termos de busca no banco de dados.



Exemplo: Imagem com legenda em latim

5. Preparação dos metadados das imagens

Inserção de Informação (Conteúdos)

Inserção das informações de interesse no banco de dados, por exemplo: tradução do título da imagem, tradução das legendas, para servirem como termos de busca no banco de dados.

Inserção no banco de dados

6. Importação para o Servidor Web

Esta etapa é realizada pela equipe de Tecnologia da Informação (TI)

Ilustra o resultado, pois na caixa de busca está escrito NAVIO, e a imagem é localizada mesmo com o título em latim *NAVIVM*.

Resultado da busca

No âmbito da preparação do acervo digital da Brasiliana USP, há vários desafios técnicos enfrentados em cada uma das etapas acima ilustradas – desde o próprio método de captação das imagens digitais, até a implementação de uma interface para o usuário que atenda satisfatoriamente suas necessidades de acesso. Para todos esses desafios, temos buscado, pesquisar caminhos e formular metodologias, com bons resultados até este momento.

Entre as áreas que estão em constante alteração e desenvolvimento está o processamento digital e a catalogação do material iconográfico (cf. 3 acima) de um modo geral (cf. 4 e 5). Neste projeto, propomos um primeiro enfrentamento desses dois desafios em conjunto, na forma de consolidação de um protocolo diferenciado de tratamento de documentos não-textuais e iconográficos e a produção de metadados para a catalogação das imagens dos textos. Em alguns casos, as imagens necessitam de um processamento digital diferenciado dos livros em geral além de apresentarem características também singulares para a descrição bibliográfica como legendas e outras informações textuais cuja correta descrição deve ser o pressuposto inicial de uma boa catalogação. Nos casos em que as imagens não apresentam informações textuais, naturalmente apenas a correta leitura do texto em que cada imagem se insere permite uma descrição global – e consequentemente, uma catalogação adequada.

Em vista disso, as atividades deste projeto foram planejadas como um trabalho de colaboração direta com a bibliotecária responsável nas etapas de extração e preparação de metadados (3,4 e 5 dos quadros acima), como se resume na sub-seção a seguir.

4.2 Plano de Trabalho

4.2.1 Capacitação técnica e teórica

O trabalho envolverá dois âmbitos de capacitação: a capacitação técnica para o uso das ferramentas computacionais relevantes ao processamento digital de imagens, e a capacitação teórica no campo da descrição documental. No primeiro âmbito, a capacitação se dará nos dois primeiros meses de bolsa, sob a coordenação da área de processamento de imagens digitais. No segundo âmbito, a capacitação será continuada, sob a coordenação da funcionária técnica de nível superior (bibliotecária) compreendendo do primeiro ao décimo mês. A capacitação teórica e técnica será efetuada por meio de treinamento individuais entre bolsista e a coordenador de área técnica sob a supervisão docente. A capacitação teórica no campo da descrição documental será efetivada por meio de reuniões quinzenais entre a coordenadora docente, a funcionária técnica de nível superior (bibliotecária) e todos os bolsistas, em reuniões quinzenais para discussão de uma bibliografia pré-selecionada relevante.

4.2.2 Trabalho de descrição dos documentos e produção de metadados

Esta etapa será coordenada por uma funcionária técnica de nível superior, bibliotecária responsável pela equipe de metadados da Biblioteca Brasiliiana Digital. O trabalho será desenvolvido em cinco etapas:

- 1) Seleção e apresentação dos conteúdos digitalizados;
- 2) Verificação do conteúdo iconográfico;
- 3) Extração das informações relacionadas aos títulos, legendas, imprensa, autoria (de livros e imagens);
- 4) Inserção destas informações nos respectivos banco de dados (Livros e Imagens);
- 5) Publicação no site da Brasiliiana Digital (<http://www.brasiliiana.usp.br>).

4.2.3 Avaliação e apresentação de resultados

Além dos encontros quinzenais em torno das leituras dos textos, o grupo terá oportunidade constante de interação, graças à concentração dos trabalhos no Laboratório da Brasiliiana. Estão programadas ainda duas jornadas de avaliação e exame dos resultados do trabalho - uma ao final da fase de treinamento, outra ao final da fase de edição. As jornadas terão uma semana de duração, tempo ao longo do qual o grupo de bolsistas e a coordenadora docente se reunirão intensivamente para elaborar relatórios internos dos progressos dos trabalhos. Ao final do ano de trabalhos, está planejado um seminário, no qual os bolsistas poderão apresentar os resultados do grupo publicamente, no ambiente do Laboratório da Brasiliiana.

5. Cronograma

ATIVIDADES	agosto	setembro	outubro	novembro	dezembro	janeiro	fevereiro	março	abril	maio	junho	julho
	X	X										
Capacitação inicial para o uso das ferramentas	X											
Seleção e apresentação dos conteúdos digitalizados e Verificação do conteúdo iconográfico		X	X									
Processamento digital das imagens				X	X	X	X	X	X	X		
Extração das informações; inserção destas informações nos respectivos banco de dados (Livros e Imagens) e publicação no site da Brasiliana Digital.				X	X	X	X	X	X	X		
Jornada de avaliação e exame dos resultados iniciais						X						
Jornada de avaliação e exame dos resultados finais											X	
Elaboração do relatório final										X	X	
Seminário de apresentação dos resultados finais												X

6. Referências Bibliográficas

BAIRD, Henry S. *Digital Libraries and Document Image Analysis*. Seventh International Conference on Document Analysis and Recognition, 2003.

MEY, Eliane Serrão Alves; SILVEIRA, Naira Christofeletti. Catalogação no plural. Brasília: Briquet de Lemos / Livros, 2009.

MINDLIN, José. Destaques da biblioteca Indisciplinada de Guita e José Mindlin. São Paulo: Edusp, 2005.

PUNTONI, Pedro. Para uma Biblioteca Brasiliana Digital. Projeto de Pesquisa sediado na BBM/USP; FAPESP, 07/597833R (Auxílio Regular à Pesquisa), 2007.

TAMMARO, Anna Maria; SALARELLI, Alberto. A biblioteca digital. Brasília: Briquet de Lemos/Livros, 2008