Brasiliana USP
500 Years of Books going Online

Fábio Kepler, Vitor Tsujigushi, Maurício Nunes, Daniela Pires, Kollontai Diniz, Carla Piazzi, Fernão Lopes, Edson Gomi, Pedro Puntoni and István Jancsó

Brasiliana Digital Library, University of Sao Paulo, Brazil

Brasiliana USP

Mindlin's Library

José Mindlin is a collector and bibliophile who, with his wife Guita and their children, assembled over the course of eighty years a library containing almost 40,000 titles. In a gesture of extraordinary generosity, he made a donation to the University of Sao Paulo (USP) comprising the books that refer to Brazil, thus the term *Brasiliana*.

▶ The Brasiliana holds nearly 15,000 titles, comprising 40,000 volumes:
  ▷ Works of Brazilian (and Portuguese) literature, travel narratives, historical and literary manuscripts (including originals and typographical proofs), periodicals, scientific and instructional texts, iconography (including plates and illustrated albums) and artistic books (engravings).

▶ Among the countless treasures, a number of works are particularly noteworthy:
  ▷ Von Martius's Flora Brasiliensis (40 volumes), together with a complete set of the naturalist's published works;
  ▷ One of the most complete collection of works by travelers and visitors to Brazil, from the 16th to the 19th centuries;
  ▷ Rare manuscripts, including one of the few known copies of Gabriel Soares de Souza's "Notícia do Brasil" (1580);
  ▷ A unique copy of the first edition of the "Relation de la Mission des indiens Kariris du Brésil", written by Father Bernard de Nantes in 1712;
  ▷ The first editions issued by the royal printing press in Brazil at the beginning of the nineteenth century;
  ▷ And collections of (extremely rare) scientific journals from the nineteenth and twentieth centuries.

Figure: *Warhaftige Historia...*, by Hans Staden, 1557.

▶ To house the Brasiliana collection, the *Guita and José Mindlin Brasiliana Library* (BBM) was created in January 2005 as a unit of the Associate Dean's Office of Culture and Extension of the University of Sao Paulo.
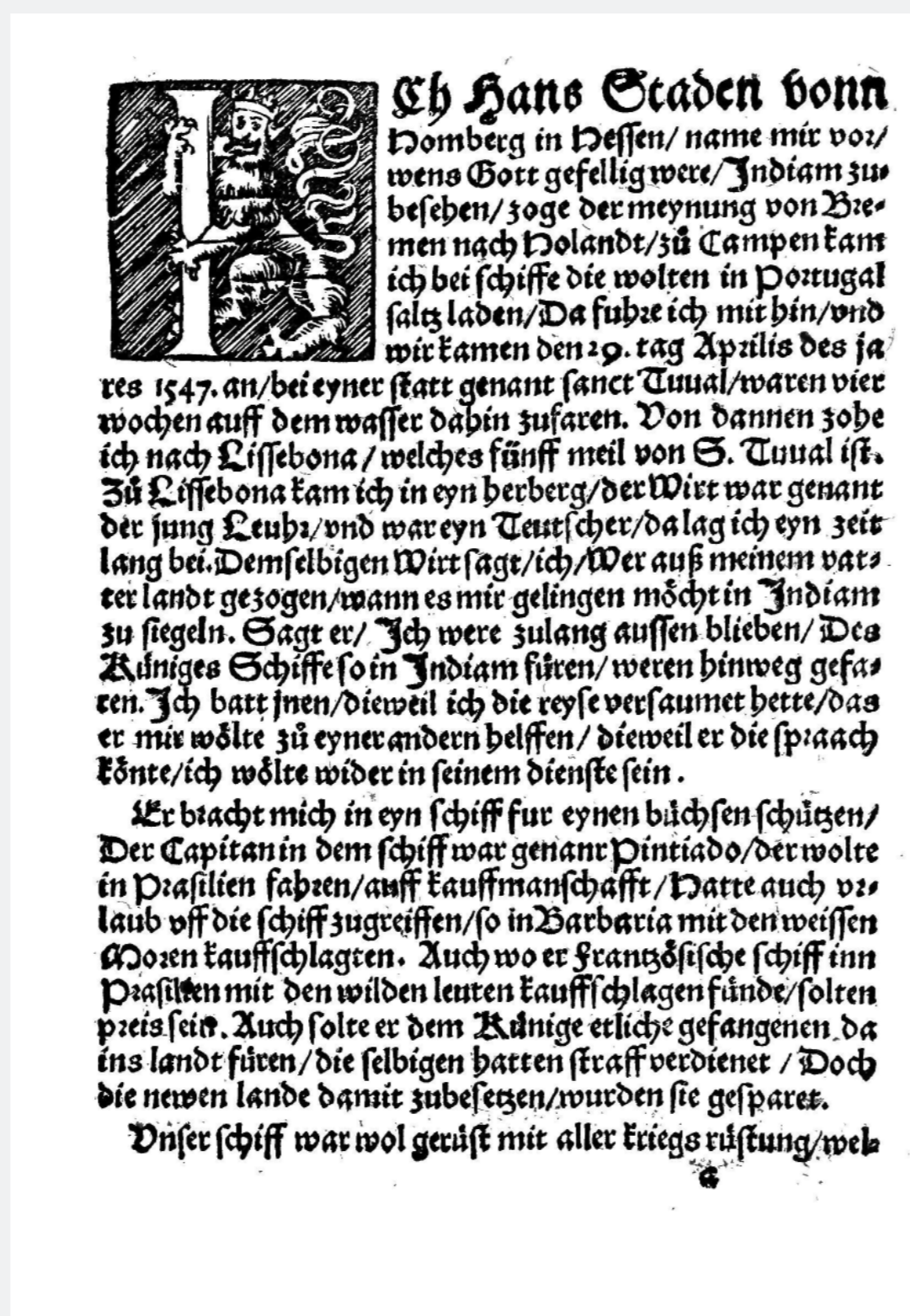
Brasiliana Digital Library

In order to provide open and broader access to the Brasiliana content, the Brasiliana Digital Library (BBD) was conceived as a project responsible for digitalizing, creating, and making available the world's largest collection of Brazilian materials managed by a university.

▶ The project was only possible thanks to the funding provided by The State of São Paulo Research Foundation (FAPESP)

▶ The BBD team is currently formed by around 40 people, working in several different fronts

▶ Most of the team works at a temporary laboratory built near BBM's construction yard

Digitalization Process

▶ The project was afforded with a Kirtas APT-2400 scanning robot.

▶ From Mindlin's library shelves to online PDF files:
  ▷ A team is responsible for selecting a book, digitalizing it with the scanner, and saving the scanned pages to a given directory on the server.
  ▷ Another team works with the scanned pages images.
    ▶ Cropping the actual page, aligning and rotating.
    ▶ Cleaning the background from artifacts, as much as possible.
    ▶ Saving in high resolution the illustrations within that book.
    ▶ Converting the book to grayscale.
  ▷ Then a team gets the processed book images and does the final touch: a PDF file is created, containing OCR text if possible.
  ▷ During this process, other teams work with the book's metadata, a review for it, and sometimes a special page for its announcement.
  ▷ When everything is done and reviewed, the book is exported from the development server and batch uploaded into the production server.

Figure: *Nova Lusitânia: História da Guerra Brasílica*, by Francisco de Brito Freire, 1675.

Going Online

▶ Brasiliana's repository is based on DSpace 1.5.2. We have done a bunch of modifications and configurations over the standard installation.
▶ The default interface is Manakin, because it allows a greater level of customization.
▶ Firstly, we needed the portal to have the looking designed by the design team. This involved not only modifications to CSS, but also some structural changes.
  ▷ The search box had to go from the right sidebar to the header, always being visible.
  ▷ Since the goal is to provide open and unrestricted access to the repository, there is no need for the login and user information box. Also, all metadata editing is done in the development server, and the items are injected in the production server via batch upload. So we tried to disable the E-Person aspect, but this raised some errors. So now we just hide the login box.
  ▷ We wanted thumbnails for items and collections.
  ▷ Many others minor customizations.
▶ We would need to frequently publish news and multimedia content in the site's main page. This led us to decide to use a CMS at the front of the repository. So currently, Drupal is being used, with links to the DSpace repository, which has backlinks to the start page. Both system have the same theme, and so from the user's perspective the navigation is smooth and transparent, although they have clear distinctions.
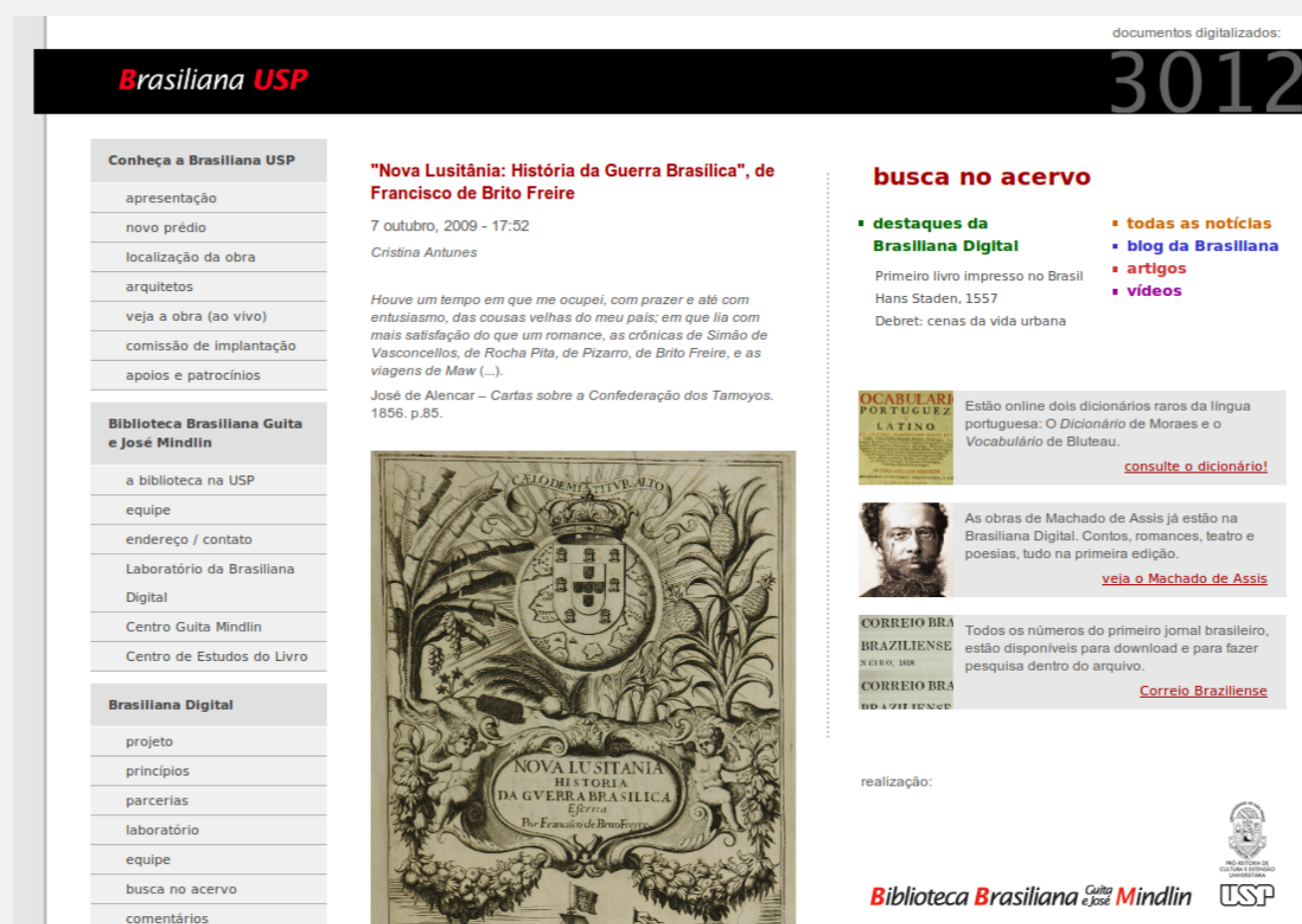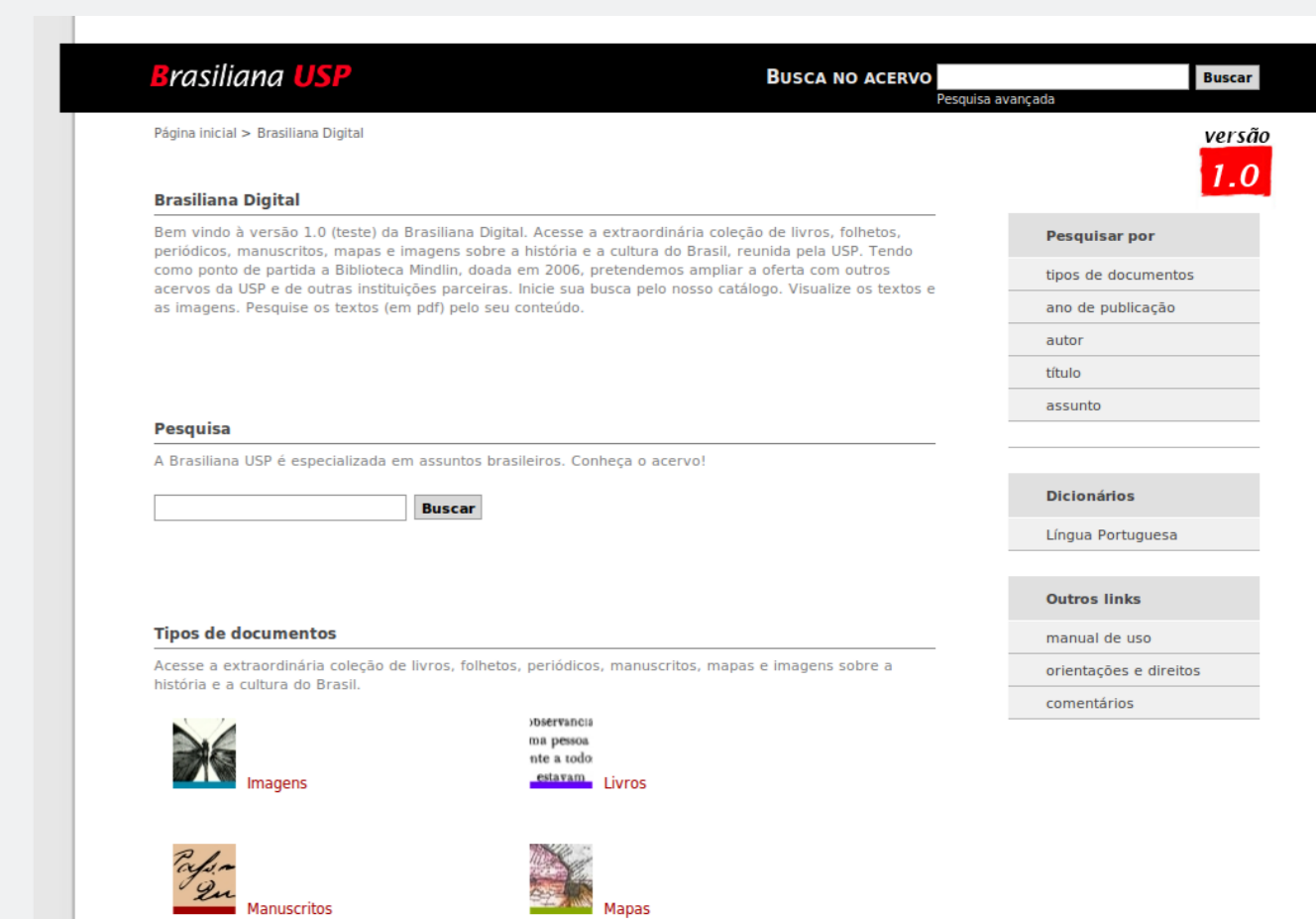

Figure: Site interface in Drupal.

Figure: Site interface in DSpace.

Future Plans

▶ The next version of the Brasiliana Digital Library is already under development. Many developments will be made at the source code level.
▶ Firstly, only one PDF document will be available, instead of the currently two (100 dpi and 300 dpi), because it will be compressed.
▶ An online PDF viewer will be developed, with functionallity added over time, such as search results highlighting, searching inside its content, and text selection over the page's image (if OCR is available).
▶ Sets of items and search results will be displayed in a grid with thumbnails an some metadata (instead of the current list interface).
▶ We will try some semantic search and new smart faceted navigation.
▶ Using more semantic information will help with our current subcollection issues
  ▷ We have many heterogeneous collections of journals and periodicals, for which we would like to provide metadata at the level of articles, as well as for issues, volumes and title.
  ▷ We have seen some solutions and suggestions, but our problem is further complicated by the fact that we do not have only this kind of item, but also many ones in collections like books, maps, manuscripts, images, etc.
  ▷ Dynamically built collections navigation seems like a solution, which we are investigating.
▶ List of Authorities and Controlled Vocabulary will be added to DSpace. Currently, they are used outside DSpace's interface.
▶ The website is currently only in Portuguese, with the exception of some metadata fields and books written in foreign languages. We plan to provide at least an English translation.
▶ A better integration with Drupal is desirable, at least at the search level.

Figure: Image no. 29, by Hans Staden, 1557.