

Calabash

Preserving & Presenting an Online Journal

By Joanna DiPasquale and Jennifer Vinopal
joanna.dipasquale@nyu.edu; jennifer.vinopal@nyu.edu

In the spring of 2003, Bobst Library received a request to help a popular NYU print journal, *Calabash: A Journal of Caribbean Arts and Letters*, “go digital.” Even in print format, *Calabash* is a “multimedia” publication, including fiction, poetry, and visual art, and representing work from around the globe. Because of its varied nature, bringing this exciting journal into the digital realm presented numerous technical hurdles: how could Bobst Library capture and digitally present this mixed media in the authors’ intended fashion; and what was the best way to make the journal available online for an audience with divergent technical capabilities and capacities? The challenges brought up by the digitization of *Calabash* have allowed Bobst Library to develop solutions, apply our knowledge to new projects, and bring the journal to the World Wide Web.

Bobst Library began the *Calabash* project with the Summer-Fall 2003 issue (Vol. 2, No. 2, <http://library.nyu.edu/calabash/toc.html>). The initial challenges were formidable. First, how could

we preserve the unique formatting of poetry in electronic format? The XHTML markup language, even with its excellent Cascading Style Sheet (CSS) capabilities, did not represent a universal solution—style classes would have to be adjusted or added per poem per issue, and some features may not be perfectly compatible across browsers and operating systems. Our goal was to protect the interest of the artist’s work while finding a web-ready format that could be easily delivered.

We also learned that Gerard Aching, editor of *Calabash*, did not necessarily want to forsake the benefits of a print journal—like turning pages—simply for a digital experience. As Gerard wrote:

“Almost a year ago, when we began to mull over this new stage of the journal’s young life, we discussed this move with a fair degree of skepticism. First and foremost on our minds was the bias that an electronic journal could not replace the beautiful objects that the first three [printed] issues of *Calabash* were and always will be.... Yet our unrepentantly romantic inclina-

tion to lament the loss of certain features that we associate with reading the journal—the tangible weight of every issue, the contemplation of the visual art that graced each cover, the indelibility of its printed words, and the reflex that our pages can only be turned—soon yielded to economic realities.”¹

To keep production costs low, an electronic format was a viable solution, but Bobst Library wanted to foster rather than thwart those “romantic inclination(s)” that made *Calabash* a wonderful print journal. In addition, because *Calabash* is international in scope—with a readership representing North America, Latin America, Africa, the Caribbean, Europe, and other areas—certain browser accessibility concerns had to be addressed.² Because of varying hardware, software, and—most importantly—connection speeds, we could not rely on special browser configuration requirements (e.g., JavaScript readability), additional technology within an Internet browser (e.g., Java applets), or the latest

1. Gerard Aching, “Preface.” *Calabash*, Vol. 2, No. 2, p. 1. Accessed on March 1, 2004, at <http://library.nyu.edu/calabash/vol2iss2/0202001.pdf>.

2. Taken from interviews with Gerard Aching, from *Calabash* print subscription statistics, and from NYU Libraries website statistics (“Visitor Domains,” March 1, 2004).

standards to deliver this information. Our documents needed to be light, quick, and portable.

Joanna DiPasquale, Web Administrator at Bobst Library, began contemplating the *Calabash* interface with these problems in mind. Knowing that cumbersome graphics on a website might hinder access, she designed small-scale graphics and made extensive use of CSS for color and overall layout. Luckily, the authors' submissions for the Summer-Fall 2003 issue were already in electronic form, and thus the primary task became finding the right electronic format in which to present the journal.

In consultation with Jerome McDonough and the Digital Library Team at Bobst Library, Adobe Acrobat's Portable Document Format (PDF) was selected. Acrobat Reader software is free and easily downloadable, and often comes packaged with browsers. It also represents a technology that takes into account many library/information technology issues: for example, Adobe and the Digital Library Federation (DLF) have examined the possibilities of archival formats and quality as the technology ages. Importantly, PDF allowed us to preserve the unique formatting of poetry, and to insert additional pages as needed.

The capacity to incorporate additional pages was especially necessary to address the visual component of *Calabash*. Delivering high-quality images from 35mm slides over the Web while trying to minimize file size presented quite a challenge. From each slide, Joanna scanned the images at 300dpi (dots per inch), and then reduced to 200dpi post-scan. She then inserted the image files into a PDF text document that contained information and interviews about the

artist, already set to 200dpi. The result: a good-quality, web-ready article that preserved and incorporated mixed media into one format.

After successful completion of a digital version of the Summer-Fall 2003 issue, we turned our attention to past issues of the journal. *Calabash* had produced three previous journals from 2001 to 2003, and we had paper copies (i.e., "hard copies") of each publication. Unfortunately, we did not have the luxury of starting with electronic copies of the articles, so we turned to the Studio for Digital Projects and Research at Bobst Library to begin the digitization project. Jennifer Vinopal, Services Manager of the Studio, led the project.

Our first challenge at the Studio

**We thus are able to
recreate in the digital
realm the user's
familiar experience of
"thumbing through"
each issue.**

lay in the physical characteristics of the journals themselves. *Calabash* was mostly text-based, but scanning produces an image of the page—not a file that contains editable words and letters. To make those page images easily readable, we needed a good quality scan that properly displayed each character on screen.

The plan was to scan the pages, save the images in the JPEG image format, then bring them into Adobe Acrobat in order to produce PDF documents. The better the quality of the scanned image, the larger the size of the electronic file produced, so we tested the process at various dpi to determine what quality scan would create highly

readable text without a file size too large for web delivery. We determined that scanning the pages in 8-bit grayscale at 200dpi met our needs.

While we were pleased with the way the page images looked, we also wanted to make the documents searchable, so readers could find words or phrases within each text. The Studio's optical character recognition (OCR) software, ABBYY FineReader, was the perfect solution. It "reads" each page image and translates what it sees into machine-readable text (i.e., a series of individual characters like those produced by a word processor).

We then turned our attention to the images of non-text items (such as pictures of studio artwork). The color photographs in the original *Calabash* publications were produced in halftone, a printing technique that uses small dots of color to create the illusion of a continuous range of tones. When halftone images are scanned, a moiré (rippled) pattern may result from the combination of the printing dots and the pixels produced in the scanning process. Bobst Library's Digital Conservation Specialist, Melitte Buchman, cautioned us about this effect and suggested that we carefully test the process of scanning and conversion into PDF to minimize the moiré effect.

In digital scanning projects, images usually are scanned at a high resolution to produce a high-quality, archival copy of the image. However, we were quite surprised to discover that the scanned halftone photographs looked better when scanned at a much lower quality than we had used for the page images themselves. After much trial and error, we found that color photographs scanned at

100dpi in 24-bit color and saved as JPEGs integrated nicely with the text images into PDF documents. Once we scanned each page of each issue, we used Adobe Acrobat to combine all the JPEG images of text pages and photos and to convert them into one PDF document per article.

Web pages for each individual issue of *Calabash* offer links to download these article PDFs, and the design and layout of the issue's web pages maintain a print-like browsability. We thus are able to recreate in the digital realm the user's familiar experience of "thumbing through" each issue. But, what the digital environment offers that the print version cannot is a "search this site" feature that gives users the ability to search the OCR'd text files "hidden behind" the PDF documents.

The PDF archives of past *Calabash* issues are forthcoming. We are adding new content and articles each day, and hope to finish by Summer 2004. To access *Calabash*, simply point your web browser to: <http://library.nyu.edu/calabash/>.

MORE ABOUT DIGITAL IMAGING AND CONSERVATION

- Conservation On-Line (CoOL) from Stanford University, at <http://palimpsest.stanford.edu>. In particular, resources on Digital Imaging at <http://palimpsest.stanford.edu/bytopic/imaging/>, are useful for learning about industry standards, conservation methods, and conversions.
- The Digital Library Team provides a clearinghouse of resources on their website, at

<http://library.nyu.edu/diglib/standards.html>.

- The Studio for Digital Projects and Research enables faculty and other NYU scholars to use digital tools and content to enhance their scholarship in the arts, humanities, and related disciplines (<http://www.nyu.edu/studio/>).
- Information about Adobe Acrobat is available at <http://www.adobe.com/products/acrobat/>.

Joanna DiPasquale is Web Administrator at NYU's Bobst Library, and Jennifer Vinopal is Services Manager at the Studio for Digital Projects and Research at Bobst Library.



The main page of the Summer-Fall 2003 issue of Calabash at <http://library.nyu.edu/calabash/>.