Building Digital Archives Collections at Northern Virginia Community College

by David Anderson

We might not think of community colleges in Virginia as institutions with deep historical roots. The Virginia Assembly established the Virginia Community College System (VCCS) in 1966 in order to fill the need for two-year college programs in the state. The fiftieth anniversary of VCCS will arrive in 2016, and that event suggests an opportunity to look back at the history of community colleges in Virginia. At the Alexandria Campus of Northern Virginia Community College, we have already begun this process by developing a digital collection to house and display historical documents from our campus archives. We have identified student publications, meeting minutes, event programs, and photographs as candidates for digitization.

In "Defining Collections in Distributed Digital Libraries," Carl Lagoze and David Fielding define a collection as a "set of criteria for selecting resources from the broader information space."[1] Essentially, collections are sets of items that meet some specific criteria of provenance and pertinence. They are commonly found in libraries, archives, museums, and other cultural institutions. Digital collections take the concept and apply it to images of items on the Web. Those items might include printed text documents, printed images, video, and audio. They might also include born-digital items. Digital collections aim to extend the reach of these items beyond their permanent homes in an archive to anyone with Internet access. Digital collections make it possible to display items online that might not get as much exposure in their

> **Done well, a digital collection should tell the story of thematically similar cultural objects....**

analog formats. Done well, a digital collection should tell the story of thematically similar cultural objects to an audience.

It is my hope that NOVA's experience will enlighten and guide other similar institutions in creating their own digital collections. While institutions with valuable treasures in their collections might have a digital collections librarian or digital initiatives librarian, no one involved in this project at NOVA had any prior experience with digital collections. Despite that, we learned by trial and error and have managed to establish an effective workflow for digitizing our documents and making them searchable and browsable on the Web.

Given the limitations of our staff, it was important to start with a project modest in scope. Digital collections can start small. We identified the most appealing items in our archives—student newspapers, Campus Council minutes, commencement programs, and photographs. These core items made up our initial collections. The student newspapers reported on events throughout the college's history and in some cases remain the only record of those events. Campus Council minutes describe the nuts-and-bolts decision-making processes that contributed to the development of the Alexandria Campus of NOVA. Commencement programs include the names of graduates and their programs as well as the names of speakers. Photographs of people and original building and site plans for the Alexandria Campus add depth and color to the collections. These items have comprised the core collection we have built to show the capabilities of digital collections on the Web.

There are some preliminary steps that institutions should take

David Anderson manages the Arlington Center Library of Northern Virginia Community College. He can be contacted at daanderson@nvcc.edu.

Digitally archived materials include these early sketches of the Alexandria Campus of Northern Virginia Community College.

Newspapers published by students at the Alexandria Campus of Northern Virginia Community College — October 24, 1975, and November 11, 1985.

before engaging in a digitization project for digital collections. Through trial and error, we learned a few:

- **Identifying items.** What should be included?
- **Software.** What kind of digital collections software will be used to display items on the Web, and what kinds of capabilities should it have? What kind of image processing software is necessary?
- **Hardware.** Should digitization be done in-house or outsourced? What kind of scanners should be purchased? How many computers and monitors will be needed for the project?
- **Staff.** How can this be done without additional staff funding or new positions?
- **Storage.** Files will need to be stored on some kind of server or other storage device. Digital files can be quite large. What quality of archival digital files should be kept?
- **Metadata.** How much metadata should be recorded for each item, and how detailed should that metadata be?
- **Intellectual property and privacy.** Is it legal to display items on the Web? Does posting items like photographs on the Web violate anyone's privacy?

We identified items to be scanned that contributed to building a historical picture of Northern Virginia Community College. We reviewed a number of digital collections software options, comparing costs, ease of use, search features, and development time, and we selected CONTENTdm from OCLC. Based on our current subscriptions with OCLC, we were able to use CONTENTdm to display up to 1,000 items or 10 GB of items on the Web for free. We purchased a copy stand and a large flatbed scanner to digitize items. We were able to use existing computers and monitors to process images. For staffing, we relied on staff volun-teers, an intern, and students to work on the digitization. We tried to make the workflow simple so that almost anyone could contribute to the scanning of items. For storage, we used a networked server to store large files as well as a portable hard drive to transfer them from one computer to another without taxing the network's resources. We are still determining what level of quality of archival digital files we will be able to keep.

As we researched this project, we saw that a robust workflow is essential to making sure that digitization, metadata creation, and the building of a website are as smooth as possible. We realized the importance of tracking each item through each step in the process—removal from the archives, digitization, moving digital files, processing digital files, adding metadata to those files, uploading them to the Web, and finally preserving them in some form on a server. Establishing a clear workflow ensures that items are not scanned or processed twice and allows staff to pick up right where they left off regardless of where they are in the workflow process.

A basic workflow for digital collections has the following components:

- **Scanning items.** A copy stand or large flatbed scanner works well.
- **File management and storage space.** Once files are scanned, it is important that they have a place to be stored. Scans of images and documents can take a great deal of space.
- **Image processing.** Files directly from the scanner will probably require straightening, cropping, conversion to PDF or other file formats, and optical character recognition (OCR) processing that allows for full-text searching of items.
- **Metadata creation.** Once files are ready to go on the Web, they need metadata. Our digital col-lections software, CONTENTdm, allows us to create metadata templates that conform to Dublin Core and other metadata standards. Fields that vary from item to item can then be filled in.
- **Upload to the Web.** CONTENTdm allows us to monitor files uploaded to the Web and approve them once they are uploaded.
- **Build menus.** CONTENTdm allows users to do searches for items, and we have built custom menus to browse to collections of items as well.

Some challenges we have faced include finding networked server space to store files, finding the most time-efficient ways to scan and process images, and determining best practices for metadata within the context of our digital collections software. CONTENTdm is widely employed for digital collections, and the option to use it for free was an attractive one; but it required some time for development and troubleshooting. We worked with CONTENTdm's customer service to deal with problems as they arose.

We are continuing to work to achieve our goal of building a stable, extensible, scalable, searchable, browsable digital collection that can serve to capture NOVA's history for years to come. We hope that our model will inspire other community colleges to build their own digital collections.

Notes

1. Carl Lagoze and David Fielding, "Defining Collections in Distributed Digital Libraries," *D-Lib Magazine* 4 (November 1998), http://www.dlib.org/dlib/november98/lagoze/11lagoze.html. **VL**