Leah Broaddus and Pam Hackbart-Dean

A tradition of access

Creating a diversity news index using OCLC's CONTENTdm

OCLC's CONTENTdm digital collection management software has been used as a platform for many interesting and timely archival projects. Morris Library Special Collections Research Center at Southern Illinois University-Carbondale (SIUC) has successfully used this platform to migrate and host digitized archival photograph collections. The center acquired administrative access to CONTENTdm through the Consortium of Academic and Research Libraries in Illinois (CARLI).

Inspired by presentations at the Midwest Archives Conference and Society of American Archivists workshops, we became increasingly interested in exploring CONTENTdm's use with different formats, such as digitized oral histories and transcribed news articles. We hoped to use it to create a campus news index documenting SIUC's history of diversity—particularly its commitment to ethnic and racial diversity. Such a project might also promote interest in other un-indexed student sources. Our challenge was to find a campus collaborator with a vested interest in the project to provide funding for the extensive hours needed to populate a meaningful, searchable index. We hope that this ongoing project will serve as a catalyst for others on our campus, and perhaps beyond.


YESTERYEAR DAILY EGYPTIAN

THE *DAILY EGYPTIAN* DIVERSITY NEWS INDEX

VIEW ONLINE COLLECTION

A Tradition of Access

Decades before other public universities gave special attention to minorities, Southern Illinois University at Carbondale already had a tradition of access. From the beginning, we opened our doors to all students, no matter what race or ethnic origin. That tradition continues through our scores of programs for recruiting and retaining minority students, our high percent of minority enrollment, and our selection of minority faculty and staff members for important administrative positions. The purpose of this project is to document SIUC's historical commitment to diversity- particularly ethnic and racial diversity. View the Online Collection.

Dick Gregory Smashes Record

Highlighting Diversity in Student News

In 2006, Dr. Seymour Bryson, Associate Chancellor for Diversity teamed up with Leah Broaddus, Archivist in the Special Collections Research Center of Morris Library and undergraduate student, Brandon Smith, to begin a project to identify articles in the *Daily Egyptian* student newspaper relating to SIUC's historic African American and Latino campus populations. This ongoing project entails surveying microfilm, locating and scanning relevant articles, and creating searchable transcripts for Web access. Currently, the database spans the years 1930-1960.

From its conception in the late 1800's, the *Daily Egyptian*, preserved on microfilm in the Morris Library Special Collections Research Center, contains a wealth of information about early student life. These early articles and images from the *Daily Egyptian* newspaper have never before been search-accessible online.

The *Daily Egyptian* continues to be published. Browse today's edition or the archives.

Screenshot of the *Daily Egyptian* Diversity News Index public entry page.

How the project came to be

Initial impetus to create a diversity news index came about by happy accident. While gathering sources for an alumni reunion display, the university archivist came upon some diversity articles from the late 1950s (by scrolling through the un-indexed microfilm of the student newspaper, *Daily Egyptian)* and sent them to the associate chancellor for diversity. The associate chancellor is an alumnus and former basketball star, so the archivist included a few highlighted articles relating

Leah Broaddus is university archivist, e-mail: lbroaddu@lib.siu.edu, and Pam Hackbart-Dean is the director of Special Collections Research Center, e-mail: phdean@lib.siu.edu, at Southern Illinois University-Carbondale

to his athletic performances. A friendly note was attached, hinting at the wealth of campus historical material that must be stored in these un-indexed papers. A few weeks later the associate chancellor telephoned to ask what it would take to initiate a project of finding more articles on diversity at SIUC and republish them for wider accessibility and research. At the end of the conversation, the associate chancellor suggested that a formal proposal be written and submitted to him for consideration.

Developing the proposal

We dropped what we were working on that afternoon and sat down to formulate a plan. Using a template the director had drawn up for another project, we took a survey of the microfilmed material, checked some dates for copyright concerns regarding digitization, telephoned the library programmer to consult about two online delivery options, and drew up a brief proposal, as requested. The proposal included 1) a description of how the goal of the project was relevant to the campus and community, 2) the software we would use for online delivery, and 3) bulleted and diagrammed estimations of the student hours needed and the amount of diversity-related materials we thought we might find within predetermined year spans.

It is important to note that preserved on microfilm in the Morris Library Special Collections Research Center, the *Daily Egyptian* contains a wealth of information about student life dating back to 1869. The early articles and images had never been searchable online. Our project aimed to make them as widely available as possible to SIUC students, faculty, administration and the general public, providing a dynamic historical and academic resource for many years to come. Therefore, the plan for this ongoing project entails surveying the microfilm, locating and scanning relevant articles, and creating searchable transcripts for Web access. These digital files are further processed with optical character recognition software (OCR) to create full-text versions of the content. The resulting digitized material is searchable online.

Student is hired, trained to do the work

The associate chancellor agreed to fund a student position to work on the project. Special Collections provided the job description, the associate chancellor interviewed and identified a student whom he felt could do the work, and then he introduced the student to us for approval. Carefully writing the job description proved very important when a project involves control that expands beyond our offices. Because we foresaw that the student would be working with just one collection, on one project, and would have to be trained for such specific work, we preferred to advertise for broader learning traits and interests rather than seeking students who already had experience with digital projects.

The job requirements included:

- attention to detail and accuracy;
- ability to exercise independent judgment;
- excellent written and oral communication skills;
- ability to take instruction and follow directions;
- hardworking, meticulous, accountable and deadline-oriented; and
- general computer skills.

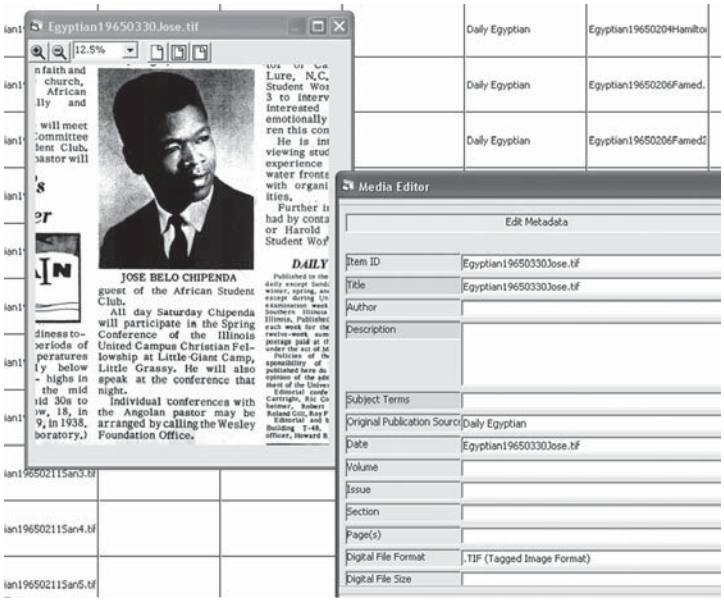
It was very simple to train the student. First he spent several weeks identifying microfilm articles, printing them off, and marking them with the source and date. Next, he patiently participated in several major false starts as we experimented with the scanning process. Following that, we trained him to edit the images in preparation for running OCR software. Then he worked on the detailed process of editing badly garbled text. He then learned to create and enter metadata describing the articles, and to upload them to the online database in CONTENTdm.

Obstacles

While the project is a success because the goal is being met, there have been obstacles along the way, mainly with scanning for OCR.

Scanning for OCR

One can expect problems when scanning from microfilm to a digital format without the



Screenshot of the CONTENTdm data entry interface.

proper equipment. We intended to run OCR software on the scans. We found that scans that did justice to the images did not do as well for text, and vice versa. Scanning the articles from the microfilm reader printouts was one option, but some of the older microfilm had been created after the articles had already faded. In many cases, the microfilm itself had also deteriorated through use, with scratches running through the text. The quality of the printouts from the microfilm reader was low, with very few contrast and lighting adjustment options.

Many of the articles would have been too large to print on a single page and required that the student apply a zoom lens view that made the text characters illegibly small on the printouts.

Finally, the quality of the images was further reduced by subsequent digital scanning. The resulting OCR from these digital images was easy to harvest rapidly, but the corrupt and garbled transcripts disproportionately slowed the editing process. The associate chancellor was revisited to discuss the possibility of outsourcing the digitization, but due to funding limitations, sending the materials off campus was not an option at that point in the project.

Micro lm versus micro che

Meanwhile the Micrographics Department on campus informed us that they had microfiche negatives of the *Daily Egyptian* on file, which were in better condition than the film. Using copies of some of the fiche, we experimented using the newest high-resolution library scanner, but even the highest resolution settings yielded dismally blurry results. The final solution was that we were permitted to train our student assistant to use the Micrographics microfiche digital scanner. The Micrographics Department asked that we work during nonpeak hours and stand ready to cede to anyone who needed the machine.

This system worked, with a few minor glitches. The scans we made were at 200 dpi initially, zoomed in as close to the borders of each article as possible at text-readable size. The quality of any illustrations embedded in the text was low, both because of the contrast settings that yielded optimal OCR results, and because of the quality of the original microfilm. However we felt that since the purpose of the project was primarily to index and describe these materials in a text-searchable manner, we would overlook the lesser quality of the article illustrations in the hopes that patrons could request better images from us directly, if needed.

Working with the software

Our consortia-provided subscription to CONTENTdm did not include access to the product's built-in PDF OCR component. There was, however, the option of running external OCR software to create and edit text documents on our own. CONTENTdm simply required that we save the text and image documents with the same file names (different extensions) and upload those simultaneously using the "Acquisition Station" software instead of working via the Internet entry-interface.

To get the database ready for the data, our library programmer set up metadata fields, including one designated for the transcript.

He trained us in mapping the fields to draw data from the files automatically and associate it into single files for each article on upload. He also provided troubleshooting when the uploads failed. The total process took several weeks of trials before running smoothly, at which point we taught it to the student, who had been occupied with scanning during that period.

Some of the image files we created would not upload, for reasons we never were able to identify. When several images failed in a batch, the program would freeze and all of the batch would fail. For the first several batches, we had to click on each individual image to import it to the acquisition station (setting aside those that failed in a separate folder). Happily, as our scanning process became more consistent, subsequent batches of images worked more smoothly.

Dick Gregory, Dizzy Gillespie Here Thursday

The SIU alumnus who became a successful comedian overnight and a famed jazz musician will be here for two shows Thursday night. Dick Gregory and Dizzy Gillespie will perform before Southern students and faculty in Shryock Auditorium at 7 and 9 p.m. shows.

The 28-year-old Gregory is a former star athlete at SIU. After several years of groping for fame with no success, Gregory hit the top with his performances at the Playboy Club in Chicago in January. He is now making the most of his position by appearing throughout the U.S.

Gregory bases his material on the racial issue. This would seem explosive on the face of the problem. However, he makes the most of the situation and leaves his audiences rolling in the aisles.

"I was so elated over this last election," declared Gregory. "This is the first time I've been far enough North to vote. You see, back in my home town we take a test: nuclear physics in Russian. And if you pass the test, they say, 'If you read Russian, you must be a Communist. You can't vote.' "

He writes his own material which he balances by lashing out at all groups, regardless of race, religion, color or national origin.

Also featured at the show will be John "Dizzy" Gillespie. He is idolized by many jazz fans and displays his genius with arrangements and showmanship. Gillespie's music represents the major trend in the American jazz movement today.

Tickets to see the two famed artists are now on sale in the Activities Development Center for $1.50.

Dick Gregory

Dizzy Gillespie

Scan of an announcement in the May 2, 1961, *Daily Egyptian*. "Dick Gregory, Dizzy Gallespie Here Thursday."

Lessons learned

The following quick tips represent some of the lessons we learned, and may be helpful to librarians contemplating small, peripheral projects involving extra-library funding:

1. First, learn the type of role an administrator wants to have in a project. Do not assume that it is the same role you would want.

2. Work quickly. When someone asks for a proposal, their enthusiasm is immediate, and if you delay, your opportunity may be absorbed or replaced by some other idea. If an administrator asks you for something, get it to him or her as quickly as you can, and then wait patiently before following up.

3. Do not mistake a campus project for a national grant—keep it simple. Be succinct, knowledgeable, and direct. Have several solutions in mind, and be clear about what is required to transform a project proposal into a reality.

4. If funding comes from sources within the university, it may be helpful to give the decision maker a choice of funding levels (for this amount, we can do A, for more we can do B, etc).

5. Expect glitches at the beginning of each stage of the project, but do not let that stop you from getting started. Adjustments will need to be made along the way when working with OCR, historical data formats, and even the most user-friendly database systems.

Thanks to a clear vision, and by being open, flexible and adaptable, we have launched a collaborative project that benefits students, faculty, and the administration of our university. Future dreams for the *Daily Egyptian* Diversity News Index include:

- wide use by students, faculty, and administration, as well as the community at large;
- feedback on improving this resource; and
- expanding the index contents and adding more formats, to include related photographs, manuscripts, and official records.

This project has already generated a lot of positive buzz. Ultimately, we have learned that by saying yes to possibilities, tapping each other's knowledge (such as expertise in scanning, CONTENTdm, etc.), embracing learning and putting snapshots of our institutional history in a position to promote ourselves, we can create something positive and satisfying for everyone. Please visit our successful *Daily Egyptian* Diversity News Index project at www.lib.siu.edu/diversitycollection.