1-1-2006

An Anniversary Opportunity: Digitization of Student Yearbooks

James Lowrey
Marquette University

Matt Blessing
Marquette University, matt.blessing@Marquette.edu

An Anniversary Opportunity: Digitization of Student Yearbooks

by James Lowrey and Matt Blessing

James Lowrey (jim.lowrey@marquette.edu) is Head of Library Systems at the Raynor Memorial Libraries at Marquette University. Matt Blessing (matt.blessing@marquette.edu) is Head of Special Collections and University Archives, also at Marquette University.

Anniversary celebrations provide archivists and librarians with many unique opportunities to build public support for their programs. Archivists, in particular, are expected to be a resource for such events. Handled adroitly, anniversary programming can offset the "dusty shelf" stereotype that frustrates many professionals. Moreover, resource allocators planning an anniversary are likely to look to their archives for ideas and special projects, initiatives that may result in additional financial and staffing resources. Describing the positive public relations value of anniversaries, archivist Tim Ericson has gone so far as to suggest "The Archivist's First Law of Outreach":

Human beings are unable to resist celebrating any anniversary divisible by twenty-five.¹

In 2005 administrators in Raynor Memorial Libraries began

considering digitization projects that might serve as a contribution to the Marquette University's upcoming 125th anniversary. Special funding was being made available to departments developing anniversary projects and related programming.

Like many comparative academic libraries, the Raynor Memorial Libraries at Marquette do not have a unit or department dedicated to digitization initiatives.² Nonetheless, archivists from the Department of Special Collections and University Archives have collaborated with staff from Library Systems to digitize parts of several important photographic collections. Staff gained valuable experience in developing these projects; they obtained training in CONTENTdm, the data-driven system selected for digital projects; researched scanning standards and taught graduate students the finer points of prepar-

ing high-resolution scans; and developed metadata fields and standards for digital collections.

Selecting a Project

Online photo collections have been described by some as "Web-site eye candy,"³ expensive digital products that some critics suggest contain nominal research value. Although those working on Marquette's digital collections disagreed with this one-size-fits-all assessment, archivists and systems librarians were interested in moving beyond item-level cataloging of photographs in order to make a larger amount of their historical collections available to online users. As Mark A. Greene and Dennis Meissner – the aforementioned critics – have persuasively argued, "good processing is done with a shovel" – not a tweezers.⁴ Greene and Meissner were writing about tra-

Raynor Memorial LIBRARIES Marquette University

Marquette Hilltop, 1915-1996

home : browse : advanced search : preferences : my favorites : help

back to results add document to favorites : add page to favorites : reference url

Search this object:

0 hit(s) :: previous hit : next hit

view: document description : go

previous page : next page

- Cover
- Introduction
- contents
- Introduction
- The Territory
 - Page 4
 - Page 5
 - Page 6
 - Page 7
 - Page 8
 - Page 9
 - Page 10
 - Page 11
 - Page 12
 - Page 13
 - Page 14
 - Page 15
- Street Life
- The Rules
- The Inhabitants
- The Games
- Hilltop Staff
- The Survivors
- Are You Ready for it
- End Materials

powered by CONTENTdm © | contact us

home : browse : advanced

	A	B	C	N
1	CDM_LVL	CDM_LVL_NAME	Title	Filename
2		1980 Hilltop	Hilltop, 1980, vol. 66	
3	0	1980 Hilltop	Cover	00000001.jpg
4	1	Introduction	frontpiece	00000002.jpg
5	1	Introduction	blank	00000003.jpg
6	1	Introduction	blank	00000004.jpg
7	1	Introduction	Page 1	00000005.jpg
8	1	contents	Page 2	00000006.jpg
9	1	Introduction	Page 3	00000007.jpg
10	1	The Territory	Page 4	00000008.jpg
11	1	The Territory	Page 5	00000009.jpg
12	1	The Territory	Page 6	00000010.jpg
13	1	The Territory	Page 7	00000011.jpg
14	1	The Territory	Page 8	00000012.jpg
15	1	The Territory	Page 9	00000013.jpg
16	1	The Territory	Page 10	00000014.jpg
17	1	The Territory	Page 11	00000015.jpg
18	1	The Territory	Page 12	00000016.jpg
19	1	The Territory	Page 13	00000017.jpg
20	1	The Territory	Page 14	00000018.jpg
21	1	The Territory	Page 15	00000019.jpg
22	1	Street Life	Page 16	00000020.jpg
23	1	Street Life	Page 17	00000021.jpg
24	1	Street Life	Page 18	00000022.jpg
25	1	Street Life	Page 19	00000023.jpg
26	1	Street Life	Page 20	00000024.jpg

Figure 1 - Rendering of table of contents via Excel

ditional arrangement and description of archival records, yet the labor intensive demands of item-level cataloging of photographs might well be viewed as the tweezers approach to digitization. Due to the sheer volume of most modern archival collections, preparing item-level metadata remains a major concern for archivists facing expanding backlogs of unprocessed records. The project team at Marquette was thus looking for a "shovel size collection" that would be appropriate for the upcoming anniversary.

Marquette's library administration and archivists considered a variety of possible digitization projects appropriate for the anniversary, including the student newspaper, the alumni maga-

zine, and the student yearbook. The decision ultimately came down to cost and complexity. Although both serials spanned more than nine decades, the cost of digitizing the semi-weekly student newspaper was four times greater than the student annual. Developers hoped that successful digitization of *The Hilltop* yearbooks⁵ (1915–1996) might leverage funding for larger scale projects in the future.

Almost as important as cost and complexity was a desire for a pilot product that would feature a high percentage of the university's 100,000 alumni. Nearly all graduating seniors had their portraits featured in *The Hilltop* yearbooks. Moreover, the student annuals provided the best coverage of

scores of extra-curricular clubs and activities. Most students were featured several times in their annual. The breadth of coverage within the student newspaper was much more limited. Especially for more recent decades, the student newspaper focused coverage on intercollegiate athletics and student government.

Preparing the Collection

Outsourcing the scanning and OCR to a vendor was a major decision.⁶ While this is a real cost, it also allows an institution to focus more on the subsequent organization and presentation of the publication, often the institution's most important contribution to a digital initiative. Unless an organization has a very well-developed infrastructure to support efficient high-quality scanning, OCR, and all the related issues, outsourcing is likely to be far cheaper, and perhaps most importantly, may serve to jump start a digital initiative. The receipt of the entire run of *The Hilltop* – totaling over 300GB for 30,000 page images with OCR transcripts – was a great incentive put the project in motion.

The CONTENTdm (CDM) image management software was the easy decision, due to our existing collections and the emergence of CDM as a standard platform for presenting image collections. In CDM parlance, a yearbook is a 'compound object'; a collection of individual items logically related into a single entity – other 'compound objects' could be postcards, photo cubes, or issues of newspaper. Once a

compound object is selected a number of decisions must be made to determine how the object is organized and presented.

First, the project team decided to retain the original structure of the individual volumes, seeking to maintain the original 'look and feel' of the volumes. In order to accomplish this, it was necessary to organize each of the 82 sets of scanned images to reflect each yearbook's sections, sub-sections, and individual page titles. Given the options available via CDM, the best method to process the images and to provide the flexibility needed for a nested structure was to define each volume with a tab-delimited file, which would then be used to import the images and transcript files and build the logical volume, or compound object. The tab-delimited files were produced from Microsoft Excel spreadsheets, with one sheet per volume, organized into decade-level workbooks.

Figure 1 depicts a sample of an Excel spreadsheet and how it renders the corresponding nested structure in CDM. The CDM_LEVEL column determines the level (9 levels are possible), and CDM_LEVEL_NAME determines when the sections break. The other key column is FILENAME, which relates the image file to the associated row of metadata. There are other columns, hidden in the illustration, which carry item-specific metadata. CDM also has a template capability to handle unchanging metadata (copyright, publisher, etc.) for every item (a single page from a book).

While the idea of preparing 82 spreadsheets to identify eve-

ry page of every volume, then reconciling each line with the images, may sound intimidating, in practice it proved very efficient. However, this approach depended upon careful organization, a thorough knowledge of what CDM expects, and expertise with Excel. Yearbooks are remarkably similar over the span of decades. The earliest volumes (1915–1935), for instance, often contained tipped-in, unpaginated color plates with tissue paper dividers. These volumes sometimes exceeded 600 pages. The 'middle' decades (1936–1965) offered much more predictable pagination. Finally, the 'experimental' decades (1966–1996) offered inconsistent pagination, seasonal supplements, sound recordings, and multiple volumes.

For every year, this structure of the volume was reflected in the Excel spreadsheet, which was determined by either using the physical volume, an annotated table of contents prepared by a paraprofessional, or, after scanning was completed, by browsing the thumbnails of the pages via imaging software while Excel was running in an adjacent window (dual monitors are ideal for this task). Using Excel fill, copy, and auto-numbering commands, as well as select functions such as MID and CONCATENATE, it was fairly easy to define the sections and their corresponding pages, replicate as needed, build the appropriate metadata text, and then check that the actual images (numbered 0000001.jpg – 0000nnn.jpg) corresponded. The major challenges were identifying unnumbered pages, plates, fold-out pages, mistakes in printing, or anything else which

interrupted 'standard' pagination. Within many volumes, for instance, the logical 'page 1' was actually the fourth image, so these volumes were defined very quickly. Most volumes required only 10–15 minutes to complete, while the more complex years required up to one hour.

Following the definition, each spreadsheet was saved as a tab-delimited text file to be used by the CDM Compound Object wizard. The wizard required the locations of the text file, the image files, and the transcript (OCR) files. It then builds the volume for review and editing, which was subsequently uploaded to our server, approved, and published. It was easy to edit the volume metadata after processing, but any substantive changes to the items (pages) or structure is best addressed by editing the spreadsheet and reloading.

This digital project also benefited from relatively few hands. One staff member performed almost all of the yearbook metadata and image reconciliation, the Excel work, and all of the uploading and approving, as well as designing and building the collection page, making the needed changes to CDM configuration, and defining the custom queries. While this may not be practical long-term, it did result in an efficient and consistent project. Given the range of styles and possible interpretation over the life of the yearbooks, this type of immersion was valuable as it became easier to predict how a volume would flow, and minimized the need to coordinate checklists, style sheets and other aids.

Figure 2 - Hilltop Entry Page. Users can access this page at: <http://digitalmarquette.cdmhost.com/HT>.

In terms of labor, all 82 volumes were defined and processed over a period of several weeks, on a part-time basis, perhaps averaging 2-4 hours per day. Concurrently, the queries, design, and layout of the page were completed after several iterations. The volumes were uploaded to our remote host via unattended FTP at a rate of 3-5 volumes per evening. A local server with a direct connection would reduce this greatly.

Evaluating the Process and the Project

The Hilltop represented a huge leap forward in both size and complexity when compared to our existing digital projects. The

small team of system librarians and archivists who planned and designed the project concentrated on getting the initiative underway and solved problems as they arose, rather than seeking to address every possible problem scenario in planning meetings. As a result, the team needed to reprocess several volumes of the yearbooks, but this seemed to be a very reasonable trade-off for the accelerated workflow. It is worth noting that the professionals who worked on the technical aspects of the project were very experienced with Excel. They had also previously built two CONTENTdm collections. This experience with the software was vital to success of the yearbook project. The 82 volumes required an estimated

60 hours of professional staff time to process, plus an estimated 20 hours for setting up the spreadsheets and managing the file archives. Preparation of the annotated tables of content by paraprofessional staff, intended as a guide for the vendor, required approximately 50 hours. Designing the graphic interface required an additional 20 hours of in-house professional labor. Finally, staff from our instructional media center assisted the project team by digitally reformatting and streaming two unique items from the university archives. A 16mm film from the mid-1930s depicted undergraduates distributing and signing yearbooks. In addition, the 1970 edition of *The Hilltop* included an 11 minute vinyl disc, a sound recording that captured some campus highlights. Adding a QuickTime movie and MP3 segment served to improve the overall attractiveness of the digital project.

Although the Raynor Memorial Libraries digitized the yearbooks primarily to provide more convenient access for alumni and university staff, the search capabilities of the digital collection also improved the efficiency and thoroughness of reference requests handled by archivists. Student employees who might previously have spent two hours scouring a decade of worth yearbooks can now compile a set of photocopies in just a few minutes. Although a relatively minor matter, requests for images found within the digital yearbooks have also increased, generating a modest source of revenue for the library.

The Hilltop project resulted in the kind of high-visibility project

that librarians had sought for the university's 125th anniversary. The digital project was completed prior the release of the campus anniversary web page. When launched, the information-rich yearbooks formed a cornerstone of this high-profile website. In addition, the campus webmaster contracted to have hyper-text links prepared to the yearbooks for more than 25,000 registered alumni within the online alumni directory. Usage statistics revealed that the digital yearbooks were browsed or searched more than 44,000 times during the initial six-month reporting period.

The Marquette University *Hilltop* can be selected from the master list of collections

at: <http://www.marquette.edu/library/MUDC>.

Notes

¹ Ericson, Timothy L., "Anniversary: A Framework Planning Public Programs," *Advocating Archives: An Introduction to Public Relations for Archivists* (Chicago: Society of American Archivists), pp. 65-81.

² The largest private university in Wisconsin, Marquette University consists of eight colleges, 35 graduate programs, and two professional schools. The Raynor Memorial Libraries has a staff of 32 librarians and 40 support staff. The Raynor Memorial Libraries house 1.4 printed volumes, while the Department

of Special Collections and University Archives provides access to nearly 20,000 cubic feet of manuscripts, photographs, and the university archives.

³ Mark A. Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *American Archivist* 68: 2 (Fall/Winter 2005). p. 248.

⁴ *Ibid.* p. 240.

⁵ Marquette University Hilltop Yearbooks, 1915-1996. Available at: <http://digitalmarquette.cdmhost.com/HT/>.

⁶ The Raynor Memorial Libraries solicited bids from various vendors and selected the firm of iArchives (<http://www.iarchives.com>) for scanning and OCR.