

CASE 6

Managing Public Affairs Records in the Digital Age

- AUTHOR:** **TIM PYATT**
University Archivist and Associate Director
Rare, Book, Manuscript, and Special Collections Library
Duke University (tim.pyatt@duke.edu)
- PAPER DATE:** April 2008
- CASE STUDY DATE:** 2004–ongoing
- ISSUE:** Public Affairs operations at institutions of higher education are rapidly abandoning paper as a medium for disseminating press releases, news stories, and campus promotional materials. Archives must develop plans to acquire and preserve these electronic records as paper surrogates are no longer produced. The Duke University Archives has found that while its “traditional” records scheduling and accessioning methods have worked to manage the intake of these records, “traditional” processing and access methodologies have not proven as effective. New workflows are needed to facilitate the transfer of electronic files and to assess file format longevity and authenticity. Re-evaluation of processing procedures and identification of appropriate access tools for electronic records need to occur.
- KEYWORDS:** Access interface, Administrative records, Appraisal issues, Data integrity issues, Metadata, Websites

Copyright by Tim Pyatt.

Institutional Context

Duke University was founded in 1838 as Union Institute Academy in Randolph County, North Carolina. By 1859 the name had changed to Trinity College and the school motto, “*Erudito et Religio*” (knowledge and religion), was adopted. In 1892 the school moved to Durham, N.C., with the support of tobacco industrialist Washington Duke. The College prospered and grew in Durham and continued to receive support from the Duke family. In 1924 James B. Duke formalized the family’s historic pattern of philanthropy with the establishment of The Duke Endowment, a charitable trust supporting universities, hospitals, orphanages, and the Methodist Church. He also donated funds to rebuild the existing Trinity campus and create a new neo-gothic campus, transforming the college into a major university. The school was renamed Duke University in honor of the Duke family. Today Duke University is a highly regarded research university with ten colleges and professional schools and nearly 13,000 students.

The University Archives traces its roots to 1892, when the Trinity College Historical Society was founded. The goal of the Society was to gather historical materials for the students to study and in the process they began to gather the records of the college. These records resided in campus attics and storage rooms as well as in the library’s Manuscripts Department until 1972, when the University Archives was established. In 2004 the archives began planning for an electronic records management program and launched its digital repository, DukeSpace, in the fall of 2006. Today the University Archives are part of the Rare Book, Manuscript, and Special Collections Library at Duke.

As the official repository for the records of Duke University, the University Archives identifies, acquires, manages, and preserves university records of enduring value, regardless of format, and makes them available for use. The archives holds more than 11,000 linear feet of administrative records, campus publications, records of student groups, and selected alumni collections and faculty papers.

Background

The University Archives staff has spent substantial time with Duke’s Division of Public Affairs staff discussing how that department manages electronic records and has worked to capture that information in record schedules. Historically, the archives has received biographical files about faculty and staff, news releases, the campus staff newspaper, and other campus promotional materials from Public Affairs. This documentation traditionally existed in paper and analog media formats. As with many campus offices, Public Affairs has started to phase out paper publications and no longer distributes press releases and news stories in the paper form. Our records manager created a retention schedule for Public Affairs, News, and Communications records, which included the following text describing electronic records:

Public affairs, news and communications offices rely almost exclusively on electronic or “born digital” records created with word processing, spreadsheet, electronic mail, website authoring, or database programs to carry out their business and activities. While the all-encompassing term “records” includes any recorded information stored on any medium, the guidelines attempt to recognize that a series may contain paper and electronic records. As electronic records are created, managed, and stored throughout their life cycle of usefulness, these guidelines should serve as a general strategy for identifying basic retention needs for different sets of information. [<http://www.lib.duke.edu/archives/rm/PubAffairs-Final.htm>]

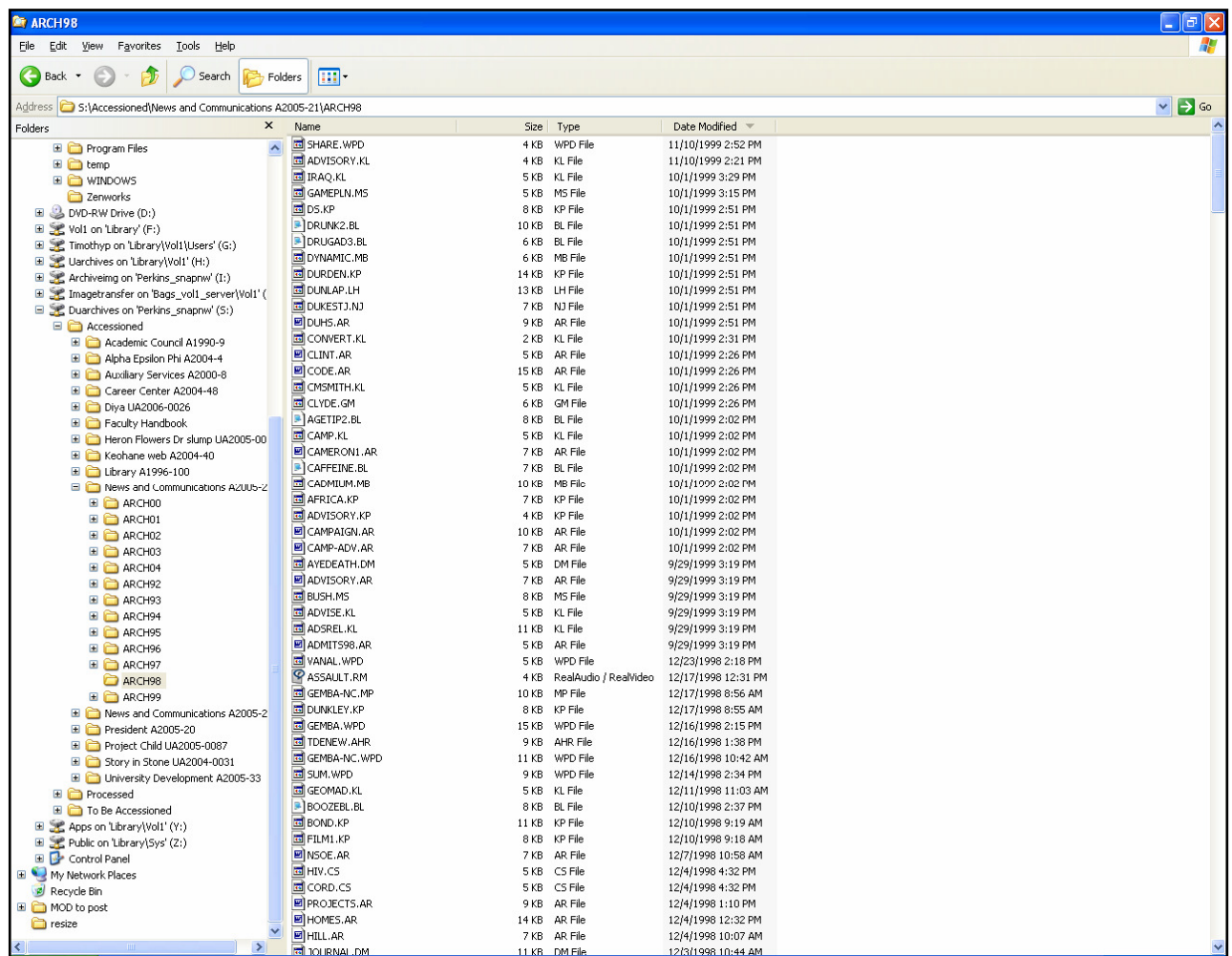
In the summer of 2004, the archives worked with Public Affairs to transfer the bulk of noncurrent paper records with archival value to the University Archives and also started discussion regarding the transfer of non-current electronic files with archival value. Public Affairs maintains the website for Duke’s president, and when President Nannerl Keohane retired in 2004, the office wanted to take her site off-line but still preserve it. Keohane, who served from 1993 to 2004, was the first Duke president to have a website. The archives acquired the files and placed them on a server dedicated for storing archival electronic records. The next significant electronic records transfer occurred in 2005 when Public Affairs migrated its website to a new content management system (CMS). Rather than migrate all press releases and news stories in electronic form to the new CMS, the office chose only to keep those from 2000 until the present “live” on the campus website. Public Affairs transferred files from 1992 to 1999, as well as source files for releases from 2000 to 2004, to the archives.

All of the Public Affairs electronic records that were received by the archives were non-current and unrestricted (open for public use). As soon as an access/delivery system could be developed, these could be made available for public use. As a test, we created PDFs of speeches by former President Keohane from her website and originally made them available on our website. While this approach works for the more than 100 Keohane speeches, we do not see it as a good solution for the more than 18,000 press releases and news stories.

Analysis

In this case, the archives acquired records through the use of FTP, discs, and some very basic web harvest. For future transfer of news releases, we have discussed using a yet-to-be-determined automated process for capturing selected content, ideally using a web capture tool that can grab content as it changes from pre-selected sites. The records transferred in 2004 and 2005 were accessioned after receipt and stored on a server in directories by accession number (see example). This parallels the process used for paper records and has been effective to date. While the metadata for the files were inconsistent (i.e., no standard naming conventions), the files were organized in directories (folders) by year and by series (i.e., all of the 1999 press releases in a single directory). Authenticity and version control was more difficult to determine as a 1999 press release may have been reused in 2003 and resaved with a 2003 date, though it still existed in the 1999 folder. This made it impossible to determine what exactly was released in 1999. Within a single year there is not a consistent filing naming convention. Each file must be opened to

determine release and content. This can be challenging as a variety of word processing software in multiple versions have been used for the files.

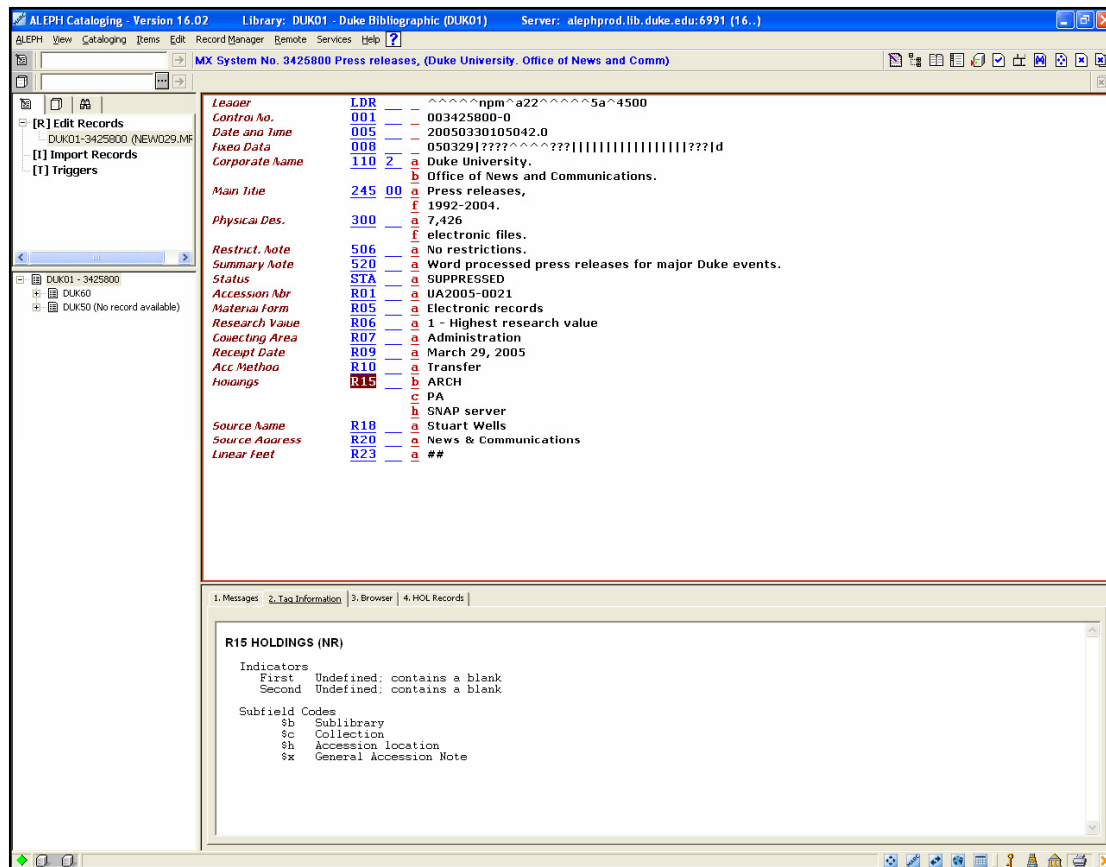


Press releases on the archives e-records server.

Access to the directories is limited to library IT staff and selected staff from archives and special collections. The server follows the library's standard back-up routine, with back-ups automatically created daily. This process is roughly analogous to our process for paper records. We accept the records, assign a unique accession number, and locate a secure storage area. The main difference is the storage area for the unprocessed records – directories on a server instead shelves in the stacks.

The process of creating the accession record itself also parallels that for paper records. The archives creates accession records using the cataloging module of Aleph, the library's integrated library system. Accession records are created as "suppressed" catalog records, using a mix of MARC and locally created fields. The administrative data of the transfer, such as office of origin, type of records, volume, and location, can still be

captured. The volume is listed as number of electronic files instead of boxes and the location is the server name rather than a stack location, but otherwise the content of the record is the same as if the files were paper-based.



Accession record for press releases

While the content and handling of the press releases, news stories, and speeches that have been transferred largely mirror the content and process for their analog predecessors, appraisal and transfer of the digital media files created by Public Affairs have proven far more difficult. These include interviews with faculty experts, Duke-produced news stories, promotional pieces, and clips associated with a news story or press release. Third-party media files are also present, such as public TV interviews with campus leaders and experts. The metadata for these records are problematic as standard naming conventions are not used and often no direct linkage to the story or press release exists. The press release, however, has an embedded link to the streaming version of the media file, giving a connection from that side of the content. Two or more versions of the media clip are often available—an uncompressed version (with a file size of over 400 MB) as well as MPEG4, Real Media, and/or QuickTime versions (file size in 40 MB range). The process for review and transfer of this content is still in the planning stages.



Example of a press release with multi-media content.

The issues involved with managing electronic files before transferring them to the archives are similar to the challenges that we face with paper-based files. Consistent use of filing names and schemes have been a source of continuous training in the realm of paper records, so it is not surprising that this continues to be an issue with digital records. The concerns about file authenticity are more vexing: in the paper era, if a 1999 press release was reworked for reissue in 2003 the original paper copy of the 1999 press release would not have been directly rewritten over the original and then re-filed with the 1999 releases. The 1999 original would have remained in the file and the “new” 2003 version would have been created and filed with 2003 releases. The ease of rewriting an existing “active” file in the digital era, even when it is the “copy of record,” raises significant authenticity issues for archivists, and these must be better communicated to the records creators.

The processing of these records, and our ability to provide access, has also been a challenge. Here, analog models do not apply as easily. For paper-based records, processing and providing access has entailed describing the records to an appropriate level, creating and encoding a finding aid, loading the finding aid and catalog record on the website and integrated library system and then making the collection available in the reading room for the archives. While processing and creating the finding aid can still be

accomplished, providing access to digital content only in the archives reading room is not the desired or expected access. For non-restricted content, users expect a web-based interface accessible from wherever they have internet access. Archival finding aids do not generally offer the same level of searchable metadata that users expect from online digital records. While the finding aid may offer searchable folder lists for paper press releases, users want to keyword search the full text of the actual items. For the archives to provide such access, each of 18,000+ press releases and news stories would have to be migrated to a standard format (such as searchable PDF), loaded on our website or another access tool, and made available through a search engine. This item-level work has been done for selected content (the previously mentioned presidential speeches), but proved too labor intensive to use for the larger volume of files.

In 2006, the archives launched a DSpace installation called “DukeSpace” to serve as the access portal for processed born digital archival files. A “University Archives” community was established. Within the University Archives community “collections” were created for administrative entities. Materials are described using Dublin Core metadata elements. Each metadata record may have one or more files associated with it and each item is assigned a persistent URL. To date our attempts to load press releases into DukeSpace have not proved satisfactory, as DSpace in its standard configuration does not scale for aggregate description and searching. A brief explanation of how the releases were loaded into the system will illustrate the challenges faced when trying to provide access large groups of related archival files. Within the “University Archives” community, a “News and Communications” collection was created. Then a “series” level metadata record was created for 1999 press releases. Since the potential existed for over 400 files to be associated with this one record, however, we did not find this to be a user friendly way to offer access to these press releases. We discovered that the individual press release titles (part of the file description) were not searchable and that full-text searching of the documents themselves was not possible with the standard configuration of DSpace. We would like to provide access without creating individual records for each press release and facilitate better search options. Archives staff members are presently exploring options to offer better access using scripts and other tools to populate DukeSpace. We are also working with campus information technology staff to enable full text search capabilities. This remains a work in progress.

The screenshot shows the DukeSpace at Duke University website in a Mozilla Firefox browser. The page displays a record for 'Press Releases, 1999' by the Duke University Office of News and Communications. The record includes a title, authors, keywords, publication date, abstract, and a URI. Below the record information, there is a table titled 'Files in This Item' listing several PDF files with their descriptions, sizes, and formats. The table has columns for File, Description, Size, and Format. The files listed are: ZEIDMAN.pdf (WASHINGTON JOURNALISTS TO DISCUSS PRESIDENTIAL CAMPAIGN), YOUTH-KL.pdf (DUKE DIVINITY SCHOOL RECEIVES GRANT), YOUTH1.pdf (TWO FALL EXHIBITS EXPLORE YOUTH CULTURE AND MEDIA INFLUENCES), YOUTH.pdf (TWO FALL EXHIBITS EXPLORE YOUTH CULTURE AND MEDIA INFLUENCES), YOHGIFT.pdf (SPIKE AND MARY YOH GIVE \$5 MILLION TO SUPPORT), and YANG.pdf (VISITING PROFESSOR AT DUKE'S FUQUA SCHOOL OF BUSINESS DIES).

File	Description	Size	Format
ZEIDMAN.pdf	WASHINGTON JOURNALISTS TO DISCUSS PRESIDENTIAL CAMPAIGN	10Kb	Adobe PDF
YOUTH-KL.pdf	DUKE DIVINITY SCHOOL RECEIVES GRANT	10Kb	Adobe PDF
YOUTH1.pdf	TWO FALL EXHIBITS EXPLORE YOUTH CULTURE AND MEDIA INFLUENCES	12Kb	Adobe PDF
YOUTH.pdf	TWO FALL EXHIBITS EXPLORE YOUTH CULTURE AND MEDIA INFLUENCES	12Kb	Adobe PDF
YOHGIFT.pdf	SPIKE AND MARY YOH GIVE \$5 MILLION TO SUPPORT	13Kb	Adobe PDF
YANG.pdf	VISITING PROFESSOR AT DUKE'S FUQUA SCHOOL OF BUSINESS DIES	8Kb	Adobe PDF

Record for 1999 press releases in DukeSpace.

Does your university archives have born-digital records?
 Share how you are effectively managing these digital records by submitting a case study to Campus Case Studies.
 Visit www.archivists.org/publications/epubs/CampusCaseStudies/.